EVOLUTION OF THE PHYTOCHROME GENE FAMILY AND ITS UTILITY FOR PHYLOGENETIC ANALYSES OF ANGIOSPERMS[1]

Sarah Mathews, Matt Lavin, and Robert A. Sharrock[2]

ABSTRACT

The phytochrome gene family encodes photoreceptor proteins that serve many functions throughout the life of a plant. From studies of the angiosperm *Arabidopsis*, the family has been modeled as comprising five loci, *PHYA–PHYE*. However, in most nonangiosperms, one locus, or at most two, is present. Moreover, it is shown here that the *Arabidopsis* model does not completely represent some angiosperm groups. For example, additional *PHY* loci related to *PHYA* and *PHYB* of *Arabidopsis* have evolved independently several times in dicot angiosperms, and monocot angiosperms (as well as *Piper*) may lack orthologs of *Arabidopsis PHYD* and *PHYE*. Nonetheless, for studies of organismal evolution, the phytochrome gene family is a potential source of phylogenetic information because the loci occur as single copy sequences, and preliminary data suggest that the various loci are evolving independently. In the plant family Fabaceae, phytochrome data are shown to provide phylogenetic resolution to a taxonomically very difficult tribe of tropical woody genera that include *Millettia*, *Lonchocarpus*, and *Derris*. In addition to nucleotide substitutions, phylogenetically informative insertions and deletions helped to resolve relationships in this group of legumes. Also, the presence of a legume-specific locus related to *PHYA* should prove to be phylogenetically informative once its taxonomic distribution is better understood.

Most molecular phylogenies of plants are inferred from one or two genes, and these usually from chloroplast or nuclear ribosomal DNA sequences. When discordance between molecular phylogenies occurs, biological phenomena such as introgressive hybridization or lineage sorting from polymorphic ancestry may explain the disparity (e.g., Harrison et al., 1987; Rieseberg & Brunsfeld, 1992; Soltis et al., 1992). Such differences also may result from lack of resolution in one of the data sets (e.g., Olmstead, 1989), or from mistaken orthology (e.g., Goodman et al., 1979; Doyle, 1992). Thus, determining organismal relationships requires that evolutionary hypotheses derived from single genes be tested with further data (e.g., Pamilo & Nei, 1988; Takahata, 1989). DNA sequences from the low copy fraction of the nuclear genome potentially provide novel phylogenetic resolution, specifically at the organismal level, since certain of the processes that lead to incongruence of species and gene trees (e.g., uniparental inheritance, nonhomologous recombination) may be less frequent.

The low copy fraction of nuclear DNA remains underexplored in phylogenetic studies of plants, and initial investigations of DNA sequences from multigene families have revealed some potential problems related to concerted evolution (sensu Zimmer et al., 1980). For example, an analysis of *rbc*S nucleotide sequences (Meagher et al., 1989) indicated that gene conversions among *rbc*S loci have occurred in each genus examined, leading to regions of "partial homology" (Patterson, 1987) and thus to the possibility of mistaken orthology. Sanderson & Doyle (1992) suggested, however, that the probability of reconstructing a reliable organismal phylogeny is high from DNA sequences of multigene families in which concerted evolution is infrequent. Preliminary data indicate that this is the case in such gene families as actin (Shah et al., 1983; Drouin & Dover, 1990; McElroy et al., 1990) and phytochrome (Sharrock & Quail, 1989; Dehesh et

[1] This work was supported by the National Science Foundation Grants DCB-9005897 (to RAS) and BSR-9118785 (to ML), and the MONTS/EPSCoR program. We thank Paul Manos, Pam Soltis, and Randy Woodson for providing DNAs of *Quercus*, *Equisetum*, and *Dianthus*, Elaine Tobin for providing tissue of *Lemna*, Jeff Doyle, Pam Soltis, and an anonymous reviewer for providing helpful comments on the manuscript, and Carol Bult and Leslie Gottlieb for providing pertinent insights.

[2] Department of Biology, Montana State University, Bozeman, Montana 59717, U.S.A.

al., 1991; Heyer & Gatz, 1992a, b; Clack et al., 1994; Adam et al., 1993); consequently, these multigene families should yield data pertinent to studies of organismal phylogenies. Furthermore, an advantage of multigene families in phylogenetic reconstruction is that, in addition to nucleotide substitution and insertion/deletion characters, the presence or absence of loci can be phylogenetically informative.

The phytochromes are photoreceptors for red and far-red light in all land plants and green algae (reviewed in Quail, 1991; Furuya, 1993). Each subunit of these large cytoplasmic receptors comprises a protein of 1100 to 1200 amino acids and a covalently attached linear tetrapyrrole chromophore. Existing in two continuously interconvertible forms, Pr, the red light-absorbing form, and Pfr, the far-red light-absorbing and biologically active form, phytochrome mediates diverse developmental responses throughout the plant's life cycle. These responses include germination, seedling hypocotyl elongation, stem cell differentiation, plastid development, flavonoid pigment synthesis, and floral induction in response to photoperiod. Modulation of plant gene expression by phytochrome is well documented (Nagy et al., 1988). While the mechanisms whereby phytochrome participates in cellular signalling remain unknown, regions of the polypeptide required for chromophore attachment, spectral integrity, biological activity, and dimerization have been identified (Cherry et al., 1993; Edgerton & Jones, 1992).

Several reports have described the presence of only a single *PHY* gene in certain nonangiosperms (Hanelt et al., 1992; Kolukisaoglu et al., 1993; Morand et al., 1993; Okamoto et al., 1993; Thümmler et al., 1992; Winands et al., 1992), while evidence of two *PHY* genes is reported for other nonangiosperms. For example, Maucher et al. (1992) refer to a putative second gene in the fern *Dryopteris filix-mas* L., although the fragment remains uncharacterized. Two unpublished *PHY* sequence fragments from *Psilotum nudum* (L.) Griseb. (GenBank accessions X74930, X74931) differ from one another in the region of overlap; and two *PHY* cDNAs from *Pinus palustris* Mill. reportedly have been cloned and partially sequenced (Furuya, 1993), while a single *PHY* cDNA from *Gingko biloba* L. is cited in the same report. However, in angiosperms, five related sequences encoding phytochrome proteins designated *PHYA–PHYE* have been characterized from *Arabidopsis thaliana* (D.C.) Schur (Sharrock & Quail, 1989; Clack et al., 1994). The genes for these five phytochromes have been mapped to *Arabidopsis* chromosomes 1, 2, 4, and 5 (unpublished), and no evidence for *PHY* pseudogenes was found. Homologs of *Arabidopsis PHYA* and *PHYB* have been characterized in other angiosperms (Adam et al., 1993; Christensen & Quail, 1989; Dehesh et al., 1991; Hershey et al., 1985; Heyer & Gatz, 1992a, b; Kay et al., 1989; Sato, 1988; Sharrock et al., 1986). A putative pseudogene most similar to *PHYA* has been reported in *Pisum* (Sato, 1990), and a cDNA clone from *Zea* containing a partial *PHY* fragment has been interpreted as a pseudogene (Christensen & Quail, 1989). Overall, these studies suggest that the gene family increases in complexity from nonangiosperms to angiosperms. This suggestion is consistent with data recently submitted to GenBank (see Results).

Nearly all *PHY* genes that are fully characterized share high sequence identity (App. 1) and structural similarity with the *Arabidopsis* loci (Fig. 1). Peptide fragments from the nonangiosperms *Psilotum* (Hanelt et al., 1992), *Anemia phyllitidis* (L.) Sw., and *Dryopteris filix-mas* (Maucher et al., 1992) share high sequence identity with the *Arabidopsis* phytochromes in their N-termini (App. 1, 2), and small internal *PHY* peptides from the alga *Mesotaenium caldariorium* (Lagerh.) Hansg. are highly similar to both N- and C-terminal peptides of other phytochromes (Morand et al., 1993). Two exceptional *PHY* genes have been described in nonangiosperms. The *PHY* gene sequence from the alga *Mougeotia scalaris* Hässel (Winands et al., 1992) contains additional introns in the N-terminal coding sequence, and in the *PHY* gene from the moss *Ceratodon purpureus* (Hedw.) Brid. the conserved N-terminal region is combined with a highly divergent C-terminal coding region (Fig. 1), which encodes a putative light-regulated protein kinase (Thümmler et al., 1992). However, in another moss, *Physcomitrella patens* (Hedw.) B.S.G., the C-terminal coding region is similar to all other *PHY* genes (Kolukisaoglu et al., 1993). No unusual *PHY* loci have been described in angiosperms.

The *PHYA–E* genes in *Arabidopsis* are differentially expressed in response to the light environment (Sharrock & Quail, 1989; Somers et al., 1991; Clack et al., 1994), and unique physiological functions have been assigned to two phytochrome proteins. Phytochrome A controls the far-red high-irradiance response (Nagatani et al., 1993; Parks & Quail, 1993; Whitelam et al., 1993), whereas phytochrome B controls red light regulation of stem length and flowering time, and the end-of-day far-red light response (Reed et al., 1993; Wester et al., 1994). This functional divergence together with high sequence divergence (approximately 50%

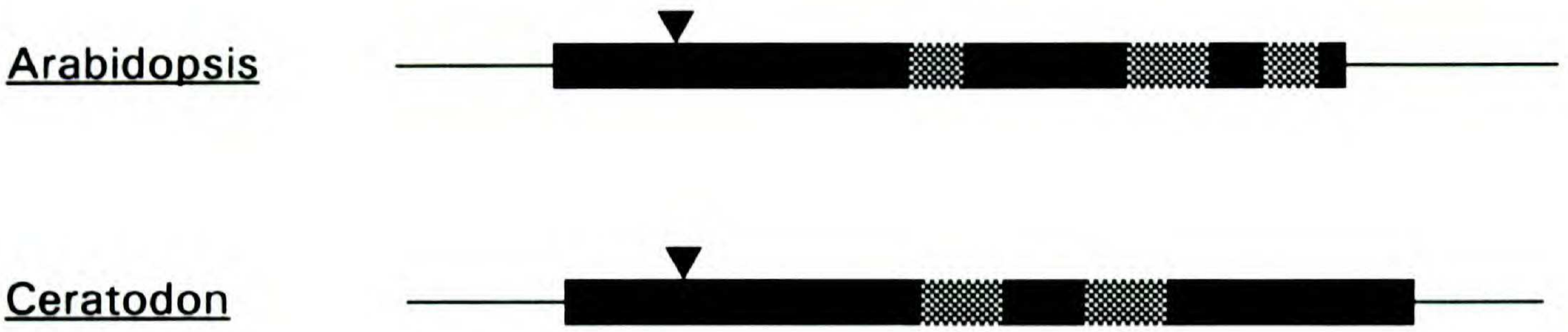

FIGURE 1. Phytochrome gene structure of *Arabidopsis* (Clack et al., 1994) and *Ceratodon* (Thümmler et al., 1992), from N-terminus (left) to C-terminus (right) showing untranslated regions (lines), exons (filled rectangles), introns (shaded rectangles), and the approximate site of chromophore attachment (triangle).

among the *PHYA*, *PHYB*, and *PHYC* loci) suggests that nonhomologous recombination is infrequent among *PHY* genes of *Arabidopsis*. If the loci are evolving independently, distinguishing orthologs from paralogs should not be difficult. To test this hypothesis, and to ascertain the phylogenetic utility of *PHY* sequence data, PCR (polymerase chain reaction) was used to sample multiple *PHY* loci from genomic DNAs of diverse species of land plants for sequence information, and these data were subjected to phylogenetic analysis.

MATERIALS AND METHODS

Total DNA was isolated from fresh, lyophilized, or dried herbarium material of taxa listed in Appendix 3 by standard methods (Doyle & Doyle, 1987). Aliquots were extracted once with phenol : chloroform-isoamyl alcohol (1 : 1 volume), and the aqueous portions were purified over sepharose CL-6B (Pharmacia, Piscataway, New Jersey) columns. To assess phytochrome diversity in early land plants, DNA sequences from different nonangiosperm phyla available in the literature (Appendix 2 and Kolukisaoglu et al., 1993) were included in the analyses with those determined during the present study. The most complete *PHY* sequence from *Psilotum* obtained from GenBank (accession X74931, lacking 510 3′ nucleotide sites out of the 3417 nucleotide sites in the full-length sequence data set) was used in phylogenetic analyses, but was not included in final alignments because it did not significantly affect the consensus sequence. Likewise, the *PHY* sequences from *Physcomitrella* and from the angiosperm *Nicotiana* (GenBank accessions X66784, L10114), were used in phylogenetic analyses, but were not included in Appendix 1. DNAs were sampled from different subclasses of angiosperms (sensu Cronquist, 1981) and, from legumes, DNAs were sampled to include two to three divergent members of the tribes Robinieae, Millettieae, and Dalbergieae in order to make preliminary evaluation of biogeographic hypotheses (e.g., Lavin & Luckow, 1993). The two species sampled from *Millettia* (*M. dura* Dunn and *M. richardiana* (Baill.) D. J. Du Puy & J. Labat) and *Sesbania* (*S. sesban* (L.) Morr. and *S. vesicaria* (Jacq.) Elliot) are not thought to be closely related within each genus.

A region of the *PHY* gene that encodes a peptide including and proximal to the chromophore attachment site was amplified using PCR, resulting in a target of 270–350 bp (App. 1). Oligonucleotides with equimolar mixtures of nucleotide pairs at two-fold degenerate sites and inosines (I) at three- to four-fold degenerate sites were designed to amplify all possible target sequences in template DNAs flanked by the conserved upstream peptide HYPATDIP (5′-CA[TC]TA[TC][TC]CIGCIACIGA[TC]AT[TCA]CC-3′) and downstream PFPLRYAC (5′-C[AG]CAIGC[AG]TAIC[GT]IA[AG]IGG[AG][AT]AIGG-3′). These peptide sequences are conserved in all *Arabidopsis* phytochromes and in the amino acid sequences inferred from other fully sequenced dicot and monocot genes, and they flank a region comprising variation likely to be phylogenetically informative. Standard PCR protocols (Perkin-Elmer, Norwalk, Connecticut) were modified to include an initial 5 cycles in which annealing temperatures were less stringent (e.g., 45–49°C). The PCR products were converted to blunt-end fragments with T4 DNA polymerase (BRL, Gaithersburg, Maryland) and were ligated to *Eco*RV-cut bacteriophage M13KRV8.2. M13KRV8.2 carries an *Eco*K cassette that facilitates screening of nonrecombinants in an *E. coli* strain which is $r_k^+ m_k^+$

(Waye et al., 1985). Transformation of *E. coli* with the ligation product yielded a population of M13*PHY* clones containing amplified genomic *PHY* sequences. Individual clones were cultured, and double-stranded phage DNA was isolated from bacterial pellets by alkaline-lysis minipreparation. Inserts cut from M13 vectors using *Eco*RI and *Hind*III were resolved on 3% NuSieve (FMC, Rockland, Maine), or 2% standard, agarose gels, and in some cases were further screened by restriction enzyme digestion to avoid sequencing duplicate clones. Single-stranded DNAs for Sanger dideoxy sequencing (Sequenase version 2.0, USB, Cleveland, Ohio) were isolated from recombinants carrying putative *PHY* inserts. In most cases, sequences of both orientations were determined, and multiple PCR products from two accessions or genera were sequenced to detect possible contamination and PCR errors. Peptide sequences were multiply aligned using ALIGN (Scientific & Education Software, State Line, Pennsylvania) and GDE 2.2 (Steven Smith and University of Illinois) and were adjusted by eye; peptide alignments were the basis for multiple nucleotide sequence alignments. For sequence comparisons, alignment gaps in certain regions of insertion/deletion were deleted, while gaps that could be identified as homologous were coded as single characters. Nonhomologous 3′ and 5′ nucleotide sites were not included in the data matrices used in cladistic and distance analyses.

Sequences were compared using maximum parsimony algorithms available in PHYLIP (Felsenstein, 1993), Hennig86 (Farris, 1988), and PAUP (Swofford, 1993). Minimal length trees resulting from heuristic search options available in either Hennig86 (mh*, bb* with no upper limit set), PHYLIP (DNAPARS), or in PAUP (CLOSEST or RANDOM data addition sequence, HOLD option set for 5 trees when applicable, STEEPEST DESCENT, MULPARS, and TBR branch swapping options activated, with branch swapping on nonminimal trees, and MAXTREES set at 10,000) were used as starting trees for further PAUP analyses (CLOSEST data addition sequence, STEEPEST DESCENT, MULPARS and TBR options activated, with branch swapping on nonminimal trees), with the latter resulting in shorter trees. Support for monophyly of clades was evaluated using bootstrap resampling (Felsenstein, 1985) and decay analysis (Bremer, 1988). Pairwise distances were estimated using the Kimura 2-parameter option available in MEGA (Kumar et al., 1993) and absolute and relative evolutionary rates were calculated by the methods of Kimura (1981) and Wu & Li (1985) respectively. All matrices subject to distance, cladistic, and rate analyses are available on request from the first author. Tree analysis and graphical output were performed with MacClade (Maddison & Maddison, 1992) and COMPONENT (Page, 1993). However, tree mapping procedures based on the model of Goodman et al. (1979), which evaluate whether incongruence of gene and species trees could be due to sampling error (Page, 1990), were not performed because of the preliminary nature of this study.

For the cladistic analysis of the full length sequences, trees were rooted by designating *PHY* sequences from *Physcomitrella*, *Selaginella*, and *Adiantum capillus-veneris* L. (Okamoto et al., 1993) as the outgroups, because they are the only fully characterized *PHY* genes from nonangiosperms. For analysis of partial sequences in angiosperms, *Selaginella* was retained as an outgroup, along with the *PHY* sequences from the gymnosperms *Gingko* and *Pseudotsuga* that were determined during this analysis.

In all cladistic analyses, first, second, and third codon positions were equally weighted for the following reasons. First, empirically determined transition/transversion ratios did not vary significantly from 1.0 for any comparisons except for between closely related legume sequences that were differentiated by very few total substitutions (e.g., ≤3% of all sites were variable). Second, results from cladistic analyses under certain differential weighting schemes are apparently the same as those from analyses under equal weighting schemes when taxonomic sampling is adequate (Albert et al., 1993; Cracraft & Helm-Bychowski, 1991). Finally, all codon positions may exhibit similar levels of homoplasy (see Chase et al., 1993); thus a rationale for excluding or differentially weighting codon positions is difficult to define. In these analyses, third codon positions, and perhaps many of the synonymous substitutions, were determined by bootstrap resampling analyses to be phylogenetically very informative, with confidence intervals for just the third codon position of between 90 and 100%, or at least as high as the values obtained for the first or second position.

Results

The orthology of fully sequenced *PHY* genes from various species to individual *PHY* loci from *Arabidopsis* has commonly been established by overall similarity (Dehesh et al., 1991; Heyer & Gatz, 1992a, b; Quail, 1991; Furuya, 1993). Similarities in gene expression and regulation have been used secondarily to imply orthology (Furuya,

1993). However, overall similarity may not reflect phylogeny, and phylogenetically related loci may differ in function due to mutations in *cis*-regulatory regions (e.g., Doyle, 1991; Li & Noll, 1994). Since orthology is best determined by shared ancestry, as evidenced by synapomorphies, cladistic analysis was used to determine the orthology of all available full length *PHY* sequences to those characterized from *Arabidopsis*. A single most parsimonious tree (Fig. 2) was generated in this analysis and it resolved the following monophyletic clades with strong (90–100%) bootstrap support: all monocot *PHYA*s, all dicot *PHYA*s, all *PHYA*s, all *PHYA*s + *Arabidopsis PHYC*, just *PHYB* and *PHYD* of *Arabidopsis*, just *PHYB*s and *Arabidopsis PHYD*, *Arabidopsis PHYE* + all *PHYB*s and *Arabidopsis PHYD*, all angiosperm *PHY*s, all angiosperm *PHY*s + *Psilotum*, and angiosperm *PHY*s + *Psilotum* + *Adiantum*. Seventy-eight trees were found by keeping all trees that were ≤30 steps longer than the most parsimonious one; all clades were retained in all trees that are 20 steps longer, except for *Arabidopsis PHYC* + all *PHYA*s. The two trees that were one step longer than the minimal length tree varied in their placement of *PHYC* as the sister group of either the *PHYA* or *PHYB/D/E* clade. These results thus suggest that, for example, the dicot and monocot *PHYA*s are orthologous, as are the dicot and rice *PHYB*s. Additionally, evidence is provided for the sister group relationship of *PHYE* with *PHYB* + *PHYD*, and for a later duplication giving rise to *Arabidopsis PHYB* or *PHYD*.

Using degenerate primers and amplification by PCR, target sequences from all five *Arabidopsis* genes, as well as from multiple *PHY* genes of other angiosperms, were recovered in single cloning experiments. Single *PHY* sequences were obtained from the nonangiosperms *Equisetum* and *Pseudotsuga* and two were obtained from *Gingko*. Inserts varied from 270 to 350 bp, and a region of insertion and deletion corresponding to residues 398 to 415 (App. 1) was eliminated from broad comparisons because nucleotide site homologies could not be determined. However, this region could be retained in narrower comparisons, where site homologies were more readily established, as in the Fabaceae data set (App. 4).

Similarly to the analysis of full-length sequences, angiosperm sequences determined in this study were cladistically analyzed to determine their orthology to the *PHY* loci of *Arabidopsis* (Figs. 3–5). Each sequence occurred in a monophyletic clade that included a single, specific *PHY* locus of *Arabidopsis*, providing evidence for distinct *PHY* subfamilies. Retention of a clade in a strict consensus tree (Figs. 2–5), resulting from the mhennig and branch-and-bound search options in Hennig86 or from heuristic options available in PAUP (see above), was considered good evidence of monophyly. Results from bootstrap resampling and decay analyses revealed that some clades were strongly supported (≥95%, d > 5–20).

The *Arabidopsis PHYA* sequence was included in a distinct monophyletic lineage in the dicot cladogram (Fig. 4). In the phylogenetic analysis of monocot sequences (Fig. 3), monocot orthologs of *PHYA* (Fig. 2) were substituted for *Arabidopsis PHYA*. Likewise, *Arabidopsis PHYA* was replaced by *Pisum PHYA* in the analysis of legume sequences (Fig. 5), also based on results depicted in Figure 2. A notable finding was that from three plant taxa, Ceratophyllaceae, Caryophyllaceae, and Fabaceae, two different PCR products were amplified that were determined to be most closely related to *Arabidopsis PHYA*. These are interpreted to be duplicated *PHYA* loci, and in legumes, the additional locus is here designated *PHYA′* (Fig. 5). These additional *PHYA*-related sequences appear to have arisen independently in the three plant groups (Figs. 4, 5). For example, the legume phytochrome phylogeny (Fig. 5) depicts this monophyletic *PHYA′* clade as being derived from within the legume *PHYA* lineage (which is thus paraphyletic). Also, it is well supported by a bootstrap value of 95%, and, in a global analysis of legume *PHYA′* with all other angiosperm loci, it is most closely related to legume *PHYA* (cladogram not shown). It thus appears that the evolution of the phytochrome gene family in the Fabaceae has involved the duplication of the *PHYA* locus. A similar argument can be made for the duplicated *PHYA* genes in Ceratophyllaceae and Caryophyllaceae (Fig. 4). In the *PHYA* subfamily, and in other cases described below, this pattern of diversification is attributed to the evolution of a new locus rather than to allelic diversity. With the exception of genes that are under frequency-dependent selection, such as alleles of the S-locus (Ioerger et al., 1990) and MHC-loci (Klein et al., 1993), levels of divergence among alleles at most loci are much lower (e.g., Gaut & Clegg, 1993; Thomas et al., 1993) than those observed among *PHYA* and the duplicated *PHYA* loci.

Sequences homologous to *Arabidopsis PHYC* were amplified commonly in monocots (Fig. 3). In dicots, only DNA of *Dianthus* yielded a sequence homologous to *Arabidopsis PHYC*. The homologs of *PHYC* in monocots were identified by their close relationship with just *Arabidopsis PHYC* in a global

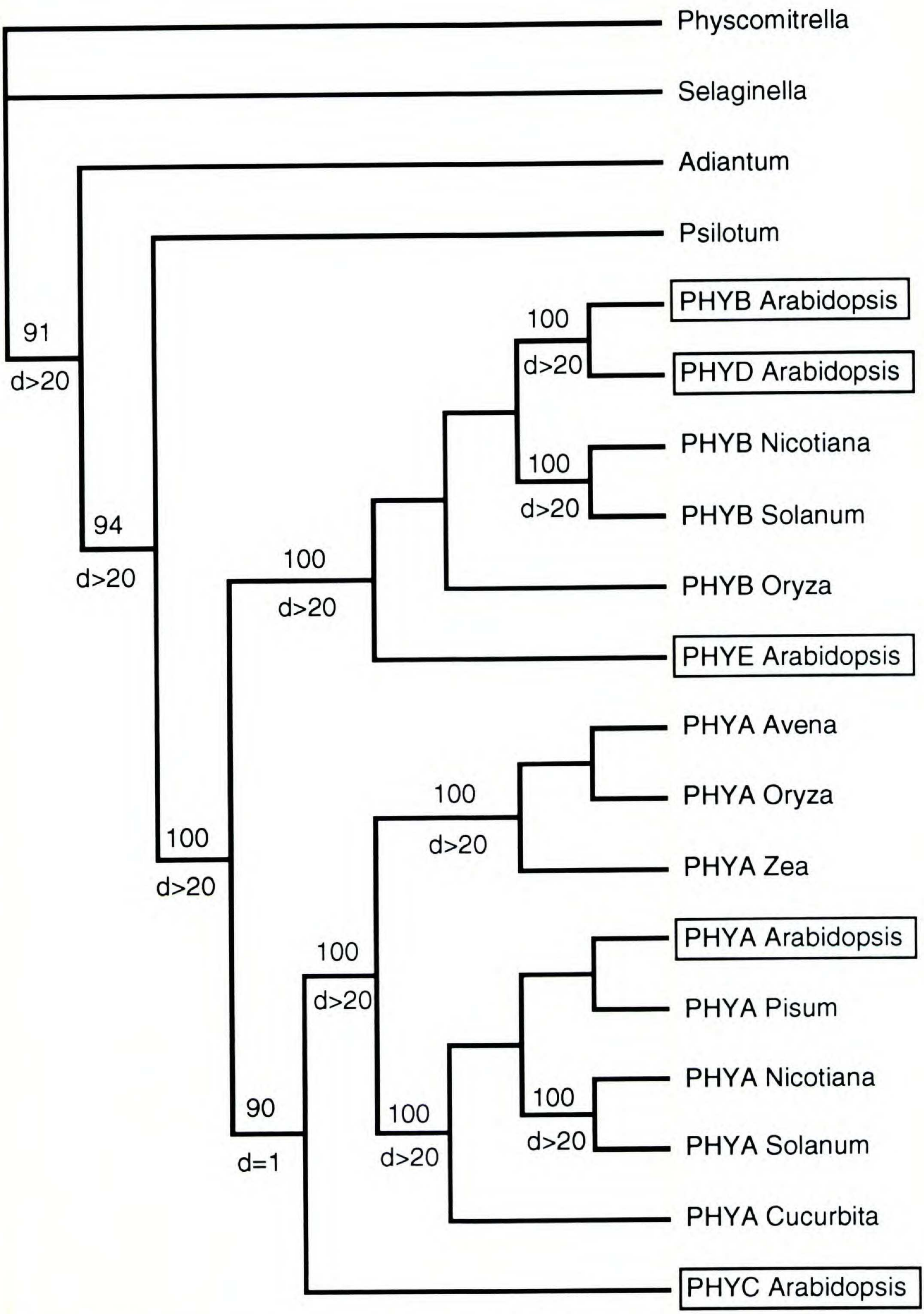

FIGURE 2. Single most parsimonious tree from analysis of 2637 variable nucleotide sites from the full-length phytochrome sequences. The length is 11,376, the CI = 0.459, and the RI = 0.502. Bootstrap values (from 500 replications) and decay indices are included on the best supported clades. The *Nicotiana* sequences were obtained from GenBank accessions X66784 and L10114.

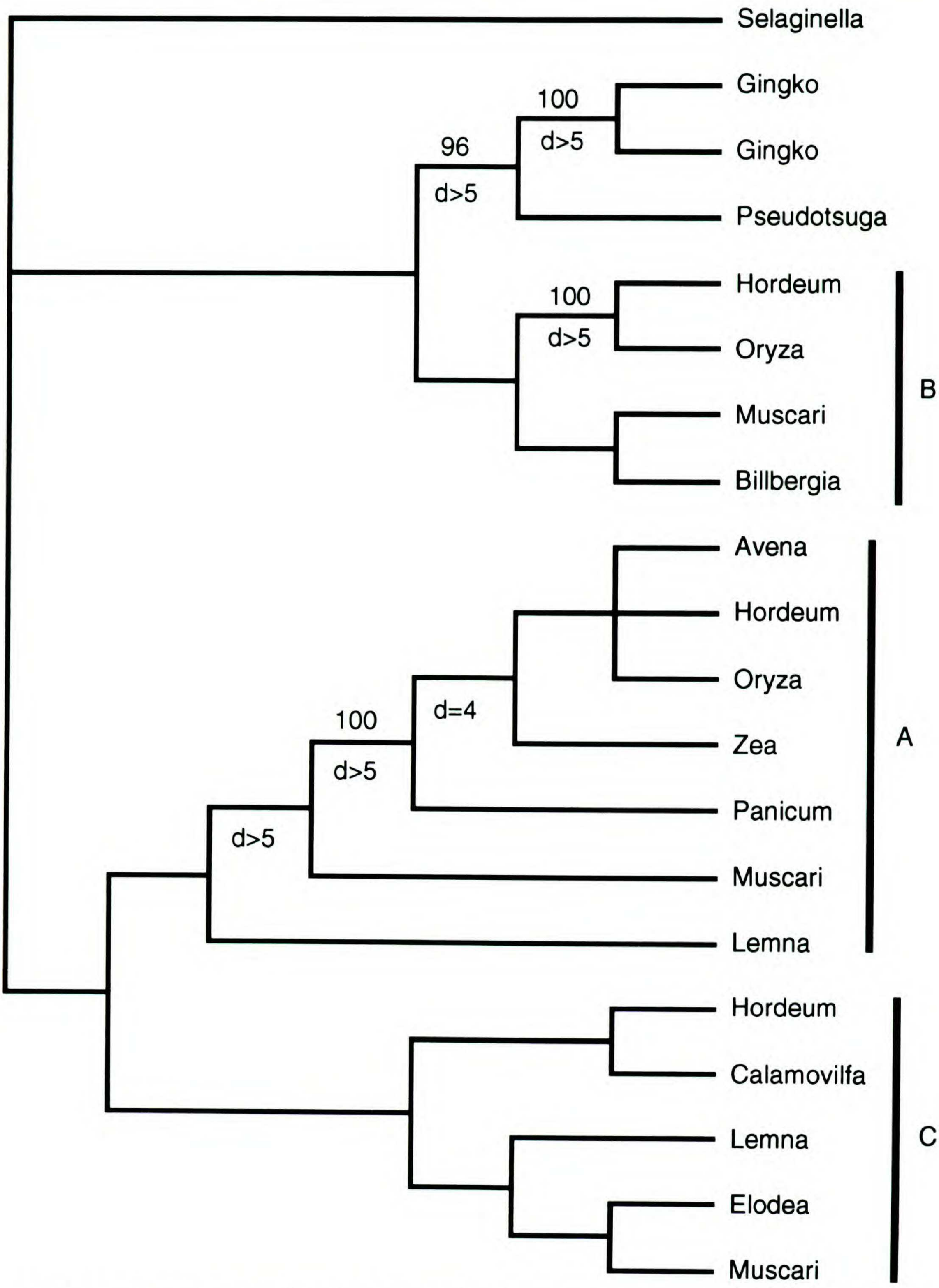

FIGURE 3. Single most parsimonious tree from analysis of all monocot sequence data, which comprised 169 informative sites. The length is 799, the CI = 0.44, and the RI = 0.52. Bootstrap values (from 500 replications) and decay indices are included on the best supported clades. Single uppercase letters to the right of the generic names are the names of the homologous *Arabidopsis PHY* loci.

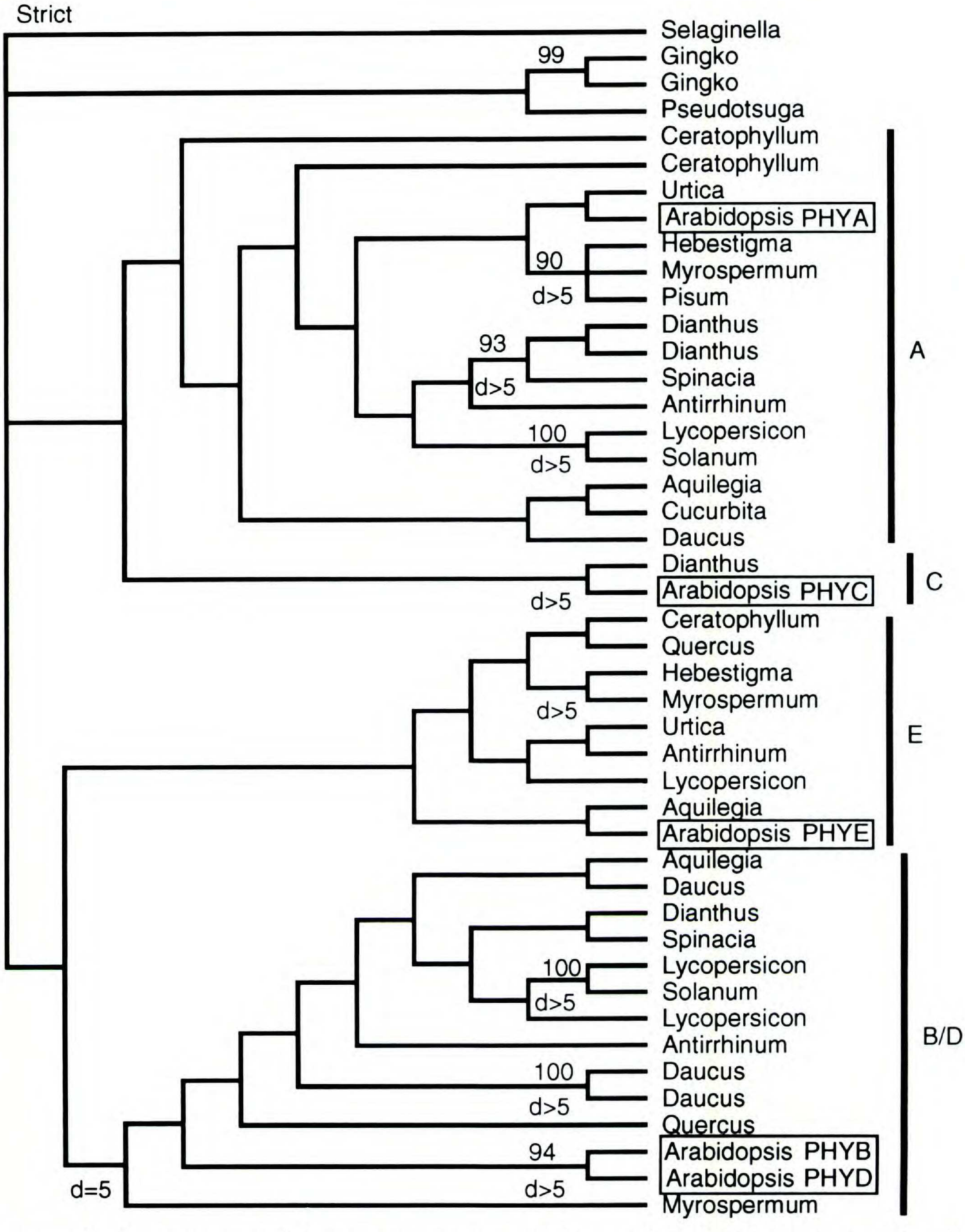

FIGURE 4. Strict consensus of four most parsimonious trees from analysis of all dicot sequence data, which comprised 172 informative sites. The length is 1743, the CI = 0.23, and the RI = 0.49. Bootstrap values (from 500 replications) and decay indices are included on the best supported clades. Single uppercase letters to the right of the generic names are the names of the homologous *Arabidopsis PHY* loci.

analysis (cladogram not shown). The *PHYC* homolog in *Dianthus* was identified by its sister group relationship with *Arabidopsis PHYC* (Fig. 4).

Sequences homologous to *Arabidopsis PHYE* were not amplified in monocots using the primer set described above. However, such homologs were commonly amplified in dicots, and the homology of these sequences to *PHYE* was readily established by the inclusion of *Arabidopsis PHYE* in monophyletic gene lineages (e.g., Fig. 4). Although the *Arabidopsis PHYE* sequence was not included in the legume data set (Fig. 5), two representative

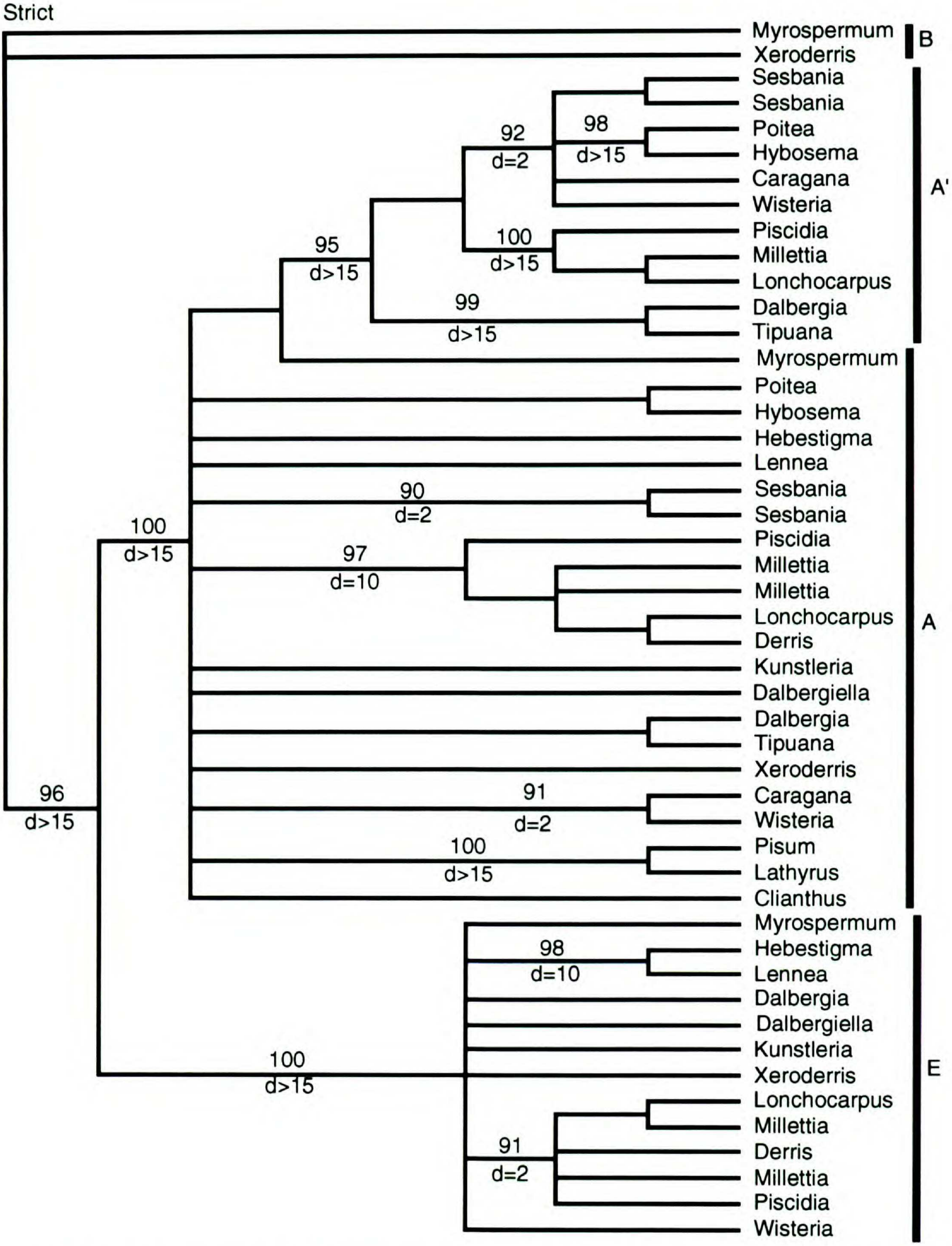

FIGURE 5. Strict consensus of 6500 minimal length trees generated from an mhennig* and branch and bound* search option on the 174 informative sites of the Fabaceae data set. Length = 545, CI = 0.534, and RI = 0.841. Bootstrap values (from 1000 replications) and decay indices are included on the best supported clades. Single uppercase letters to the right of the generic names represent the orthologs of the *Arabidopsis PHY* loci.

legumes were included in the dicot analysis shown in Figure 4, and these were part of the monophyletic gene lineage that included *Arabidopsis PHYE*. In the legume gene phylogeny, the bootstrap value for the *PHYE* clade was 100%, thus revealing how strongly this lineage is supported by the data in narrow comparisons at the taxonomic level of the family.

The evolution of genes related to *Arabidopsis PHYB* has been more complex, with the apparently independent duplication and divergence of *PHYB*-related genes in some dicot lineages, but perhaps not in monocot lineages (Figs. 3, 4). The notable pattern here is that the *Arabidopsis PHYB* and *PHYD* sequences are sister groups in comparisons including dicots (Figs. 2, 4), and together with the sequence from *Myrospermum* are the sister group of the other *PHYB/PHYD*-related sequences. Note that two *PHYB/D*-related sequences occur in *Lycopersicon*, forming a monophyletic clade, with a *PHYB*-related sequence from *Solanum*, that is separate from the clade containing *Arabidopsis PHYB* and *PHYD*; two of the *PHYB/D*-related sequences from *Daucus* also form a monophyletic clade (Fig. 4). This pattern could result from nonhomologous recombination between loci, but the hypothesis of recent divergence is consistent with the putative absence of additional *PHYB*-like sequences from monocots. Additionally, *PHYD* in *Arabidopsis* is apparently functionally distinct, as evidenced by its failure to compensate for the loss of *PHYB* function in *phyB* null mutants of *Arabidopsis* (Reed et al., 1993; Wester et al., 1994).

In the two trees with dicots (e.g., Figs. 2, 4), *PHYE* is the sister group to the *PHYB/PHYD* clade. Since *PHYD* and *PHYE* have not been amplified from monocots, the diversification of this part of the phytochrome gene family may have taken place only during the diversification of dicots. Further sampling from Nymphaeales, Piperales, Winterales, Laurales, and Magnoliales should address the question of whether the presence of just *PHYA*, *PHYB*, and *PHYC* is the ancestral condition in angiosperms. Notably, however, preliminary analysis of three sequences from *Piper* recently submitted to GenBank (Kolukisaoglu et al., unpublished), derived using a different primer pair, suggests that they are orthologs of *Arabidopsis PHYA*, *PHYB*, and *PHYC*. Alternatively, the inability to amplify *PHYD* and *PHYE* from monocots (and *Piper*) could mean that the oligonucleotide primers designed in recent studies do not recognize and amplify all *PHYD* and *PHYE* homologs. This alternative explanation should be evaluated in subsequent studies of the phytochrome gene family in monocots and in magnoliids with uniaperturate pollen.

The relationships among the angiosperm and nonangiosperm *PHY* lineages were evaluated in two additional types of analysis: (1) parsimony analyses of nucleotide sites homologous to the PCR target fragment, including all angiosperm *PHYA–E* paralogs and all nonangiosperm *PHY* sequences for which there were corresponding nucleotide data (about 330 bp); and (2) parsimony and distance analyses of amino acid sites homologous to the *Mougeotia* fragment (App. 2, about 300 amino acids), including from angiosperms shown in Appendix 1, and from gymnosperms determined in this study (with sites coded as missing). Patterns that emerged from the nucleotide sequence analyses included: (1) *Ceratodon*, *Physcomitrella*, *Selaginella*, *Equisetum*, *Gingko*, and *Pseudotsuga* most commonly occurred as sister groups of a *PHYB/D/E* clade; (2) *Mougeotia*, when not designated as the outgroup, was the sister lineage of the *PHYC* clade; (3) *PHYC*s were basal and paraphyletic in many cladograms rooted at *Mougeotia*; in others, or if *Mougeotia* was removed from analyses, a *PHYB/D + PHYE* clade and a *PHYA + PHYC* clade were most often resolved. Results of the amino acid sequence analyses indicate a major split between *PHYA + PHYC* and the *PHYB/D/E* clade, each with a set of nonangiosperms as sister group. The only common element of the two sets of analyses was the close relationship between the sequences available from *Gingko* and *Pseudotsuga* and the *PHYB/D/E* clade. Further, the robustness to perturbation of the data, which is found in the analysis of the full-length sequence data set (Fig. 2), is lost in these broad comparisons when the number of sites is limited.

Recently, single phytochrome sequence fragments (561–654 bp) from a number of nonangiosperms, including *Gnetum* and *Ephedra*, were deposited in GenBank (Kolukisaoglu et al., unpublished), bringing the total number of nonangiosperm homologous *PHY* sequence fragments available for nucleotide analysis to 15. Preliminary analyses of these sequences indicate that the data are still too fragmentary to draw conclusions regarding evolution of specific loci, especially as some of the nonangiosperm taxa represented by a single sequence are likely to have more than one *PHY* gene. Furthermore, organismal relationships depicted in these cladograms and neighbor-joining trees, except for the pairs *Ceratodon + Funaria* (both mosses) and *Metasequoia + Picea* (both conifers), are not well supported in bootstrap analysis.

Discussion

PHYTOCHROME EVOLUTION

The evolutionary pattern that emerges from phytochrome gene studies is that *PHY* gene diversity appears to be limited in nonangiosperms, where often a single gene is found, while diversity is much greater in angiosperms, where orthologs of the *PHYA*, *PHYB*, *PHYC*, *PHYD*, and *PHYE* genes discovered in *Arabidopsis* are present (Figs. 2–4). The data suggest that divergence of at least two, and most likely three, of the loci found in angiosperms preceded the diversification of flowering plants. For example, orthologs of *Arabidopsis PHYA*, *PHYB/D*, and *PHYC* have been detected in most angiosperm subclasses, and there is evidence for two loci in some nonangiosperm groups. Moreover, the model of a five member phytochrome gene family developed for *Arabidopsis* is probably not completely appropriate for all angiosperms. For example, though the PCR primers developed in this study annealed to and amplified dicot orthologs of the *Arabidopsis PHYA*, *PHYB/D*, *PHYC*, and *PHYE*, they annealed and amplified only three paralogs in monocots, *PHYA*, *PHYB/D*, and *PHYC*. The same primers applied to DNAs from the Fabaceae most commonly amplified *PHYA*, *PHYA'*, and *PHYE*; rarely did they amplify *PHYB/D* homologs, and they have yet to amplify *PHYC*. It is very possible that sequence divergence at the primer sites precludes the amplification of all loci present in some genomes, or that bias toward certain gene family members has occurred during amplification cycles; i.e., PCR selection or drift (sensu Wagner et al., 1994) has occurred. However, preliminary results indicate that the same loci are obtained from genera in Fabaceae when primers differing in GC content are used (Lavin, unpublished); likewise, certain variations of initial amplification conditions have not altered the set of loci detected in other angiosperms (Mathews, unpublished). Thus, it is likely that all five genes characterized from *Arabidopsis* did not precede the early diversification of angiosperms. Indeed, data presented here showing independent evolution of multiple *PHYB/D*-related sequences in *Arabidopsis*, *Lycopersicon*, and *Daucus* indicate that the divergence of the *PHYB* and *PHYD* loci in *Arabidopsis* occurred sometime well after the diversification of dilleniid families. Recent diversification of the phytochrome gene family in angiosperms is also suggested by the occurrence of *PHYA*-related sequences that have independently evolved in Ceratophyllaceae, Caryophyllaceae, and Fabaceae (see Fig. 5).

TEMPO OF SEQUENCE EVOLUTION

Using the 2-parameter model of Kimura (1981) to estimate distances among all pairs of full-length coding sequences, and a divergence time for *Selaginella* of 300 million years (Ma) ago (Townrow, 1968), the estimated overall rate of evolution of *PHY* lineages is 0.9 to 1.5 $\times$ 10^{-9} substitutions per site per year, or about ten times as fast as *rbc*L (Chase et al., 1993). In contrast, the rate of Jukes-Cantor corrected synonymous substitutions (K_S) among *PHY* sequences from pooid and panicoid grasses, with an estimated divergence time of 50 Ma (Doebley et al., 1990), and among tropical woody tribes of Fabaceae, with an estimated divergence time of 40 Ma (Herendeen, 1992; Wheeler & Baas, 1992) is four to five times as fast as *rbc*L (Zurawski et al., 1984; Doebley et al., 1990), or about 3.7 to 6.1 $\times$ 10^{-9} substitutions per site per year. Rates of Jukes-Cantor corrected nonsynonymous substitutions (K_A) estimated from pairwise comparisons with *Selaginella* for different portions of full-length phytochrome molecules (App. 1) indicate that the 594 bp including and proximal to the chromophore attachment site is the most conserved portion of the molecule (K_A = 3.2 to 4.6 $\times$ 10^{-10} subst./site/year), followed by the 2400 bp encoding the N-terminus (K_A = 4.0 to 5.4 $\times$ 10^{-10} subst./site/year), followed by 3384 bp comprising nearly the complete coding region (K_A = 4.3 to 6.2 $\times$ 10^{-10} subst./site/year). It is notable that K_S is consistently greater than K_A, even among the most closely related *PHY* loci (e.g., *Arabidopsis PHYB* and *PHYD*, and Fabaceae *PHYA* and *PHYA'*). The opposite pattern of substitution among codons associated with functional divergence has been used to suggest recent positive selection for divergent function among alleles (Nei & Hughes, 1991) and closely related loci (Ngai et al., 1993). However, the *PHY* loci might not be amenable to this comparison because of their more ancient divergence. Furthermore, the test cannot be precisely applied without more specific knowledge about codons associated with divergent functions.

In 42 relative rate tests (Wu & Li, 1985) used to evaluate the hypothesis that rates within and among the *PHY* loci are clocklike, 11 rate differences were significantly different ($P < 0.05$ or 0.01), given a model of rate constancy. All of these significant differences were among, rather than within, *PHY* lineages (Appendix 5), and are thus unlikely to be the source of spurious long-branch attractions in organismal phylogenies (Hendy & Penny, 1989).

IMPLICATIONS FOR ORGANISMAL PHYLOGENETIC ANALYSIS

Phytochrome sequence data is providing a high degree of phylogenetic resolution within the plant family Fabaceae, and this suggests that the phytochrome gene family, at the least, should be a promising source of data below the familial taxonomic level. Among other sorts of promising taxonomic characters is the presence of a novel legume locus related to *PHYA* (here referred to as *PHYA'*), which should eventually serve as a phylogenetic marker for a major subgroup of Fabaceae, or possibly among related families, once its taxonomic distribution becomes better known. One example of the significant phylogenetic implications that have been revealed so far is outlined below.

The phylogenetic relationships of the tropical woody papilionoid legume genera *Millettia*, *Lonchocarpus*, *Derris*, and putative close relatives of the tribe Millettieae remain poorly resolved (Polhill, 1994), despite recent comprehensive taxonomic studies (e.g., Evans et al., 1985; Geesink, 1981, 1984). *Millettia* is traditionally characterized only by its elastically dehiscent legume (Dunn, 1912), but the paraphyletic (and perhaps polyphyletic) nature of the genus has recently been confirmed by chloroplast DNA data (Liston, 1992). *Lonchocarpus* and *Derris* have indehiscent legumes; the former is traditionally distinguished by its wingless pods and a staminal tube with basal fenestrae, whereas *Derris* is traditionally characterized by winged pods and staminal tube lacking basal fenestrae (Geesink, 1981, 1984). However, these traditional characterizations have recently been disputed (Sousa & de Sousa, 1981; Sousa & Delgado, 1993). They argue that *Lonchocarpus* and *Derris* and relatives should be excluded from a close relationship with *Millettia* and allies, and placed closer to the genera of the tribe Dalbergieae, because of their indehiscent pods and putative indeterminate inflorescences. They also consider *Millettia* and close relatives to be part of the tribe Robinieae. In contrast, Polhill (1971, 1981) placed *Millettia*, *Lonchocarpus*, *Derris*, and close relatives together as a tribe separate from Dalbergieae and Robinieae (Polhill, 1981), because the three lineages have a similar phytochemistry and inflorescence structure (e.g., the pseudoracemose inflorescence).

Phytochrome sequence data from *PHYA*, *PHYA'*, and *PHYE* in these tropical woody papilionoid genera show much promise in providing at least some phylogenetic resolution to this group. The representatives of *Millettia*, *Lonchocarpus*, *Derris*, and certain allied genera (e.g., *Piscidia*) used in this analysis are consistently monophyletic in all minimal-length trees and in all three gene phylogenies (Fig. 5). Bootstrap confidence intervals above 90% in each individual gene phylogeny, and an amino acid deletion at position 405 (App. 4) in the *PHYA'* sequence, further support the monophyly of these genera. The phytochrome data suggest that this group is distinct from Dalbergieae (represented by *Dalbergia* and *Tipuana*), Robinieae (represented by *Sesbania*, *Hebestigma*, *Hybosema*, and *Lennea*), and certain other genera of Millettieae (e.g., *Kunstleria* and *Dalbergiella*). Such a grouping of *Millettia*, *Lonchocarpus*, *Derris*, and *Piscidia* (and presumably certain other genera when sampled) is consistent with chloroplast DNA data (Lavin, unpublished; see also Doyle & Doyle, 1993) and certain morphological data (Polhill, 1971). For example, this generic group is distinguished from other genera in the same tribe (such as *Kunstleria* and *Dalbergiella*), as well as the tribes Dalbergieae and Robinieae, by an inflorescence in which the flowers are fascicled along the raceme rachis, and by flowers in which the standard petals have claws that are abruptly contracted and subtended by calluses and inflexed auricles. This grouping is not consistent with whether the pods are dehiscent or not, or what type of nonprotein amino acid is accumulated in seed. That three different phytochrome loci, which are presumably under different evolutionary constraints, all reveal this same monophyletic group suggests that phytochrome sequence data will have a bearing on revealing those morphological characters that may best serve as phylogenetic markers in this taxonomically complex group of papilionoid legumes.

FUTURE DIRECTIONS

Phytochrome DNA sequence data, readily obtainable using PCR, are shown here to be informative regarding questions of organismal phylogeny in narrow comparisons, such as among closely related genera. However, the degree of resolution depicted in Figure 2 is promising for their use (if more nucleotide sites are included) in broader comparisons as well; notably, the branching order (except for the placement of *Psilotum*) is consistent with current hypotheses of plant phylogeny (summarized in Donoghue, 1994). Further, equally promising is the potential to use composite trees inferred from pairs of phytochrome loci that diverged prior to the diversification of angiosperms to determine evolutionary relationships among the

major angiosperm lineages in the manner Iwabe et al. (1989) inferred relationships among archaebacteria, eubacteria, and eukaryotes.

The data presented also raise intriguing questions concerning the evolution of individual phytochrome loci. For example, do monocots and a certain subgroup of magnoliids with uniaperturate pollen have only *PHYA*, *PHYB*, and *PHYC*, whereas in eudicots and another subgroup of magnoliids, diversification of the phytochrome gene family is much greater? If so, the *Arabidopsis* model is not completely applicable to monocots. As with the *PHYA'* locus in Fabaceae, the taxonomic distribution of *PHY* genes in monocots should provide phylogenetic insight into the divergence of monocots from dicots. Additionally, further phytochrome data, especially from nonangiosperms, potentially will reveal the history of phytochrome gene duplication events in the context of green plant phylogeny.

Exploration of such questions may be facilitated by a variety of tools; for example, preliminary data indicate that development of locus-specific PCR primers will be productive. So far, exclusively *PHYB*-related sequences have been determined from *Arabidopsis*, *Daucus*, *Quercus*, and *Spinacia* using a 3′ *PHYB/D/E*-specific primer in combination with the conserved 5′ primer.

Sequences determined in this study from taxa other than Fabaceae are available from GenBank under accession numbers UO8142–8184.

Literature Cited

Adam, E., M. Deak, S. Kay, N. H. Chua & F. Nagy. 1993. Sequence of tobacco (*Nicotiana tabacum*) gene coding for type A phytochrome. Pl. Physiol. 101: 1407–1408.

Albert, V. A., M. W. Chase & B. D. Mishler. 1993. Character-state weighting for cladistic analysis of protein-coding DNA sequences. Ann. Missouri Bot. Gard. 80: 752–766.

Bremer, K. 1988. The limits of amino acid sequence data in angiosperm phylogenetic reconstruction. Evolution 42: 795–803.

Chase, M. W., D. E. Soltis, R. G. Olmstead, D. Morgan, D. H. Les, B. D. Mishler, M. R. Duvall, R. A. Price, H. G. Hills, Y.-L. Qiu, K. A. Kron, J. H. Rettig, E. Conti, J. D. Palmer, J. R. Manhart, K. J. Sytsma, H. J. Michaels, W. J. Kress, K. G. Karol, W. D. Clark, M. Hedrén, B. S. Gaut, R. K. Jansen, K.-J. Kim, C. F. Wimpee, J. F. Smith, G. R. Furnier, S. H. Strauss, Q.-Y. Xiang, G. M. Plunkett, P. S. Soltis, S. Swensen, S. E. Williams, P. A. Gadek, C. J. Quinn, L. E. Eguiarte, E. Golenberg, G. H. Learn, Jr., S. W. Graham, S. C. H. Barrett, S. Dayanandan & V. A. Albert. 1993. Phylogenetics of seed plants: An analysis of nucleotide sequences from the plastid gene *rbc*L. Ann. Missouri Bot. Gard. 80: 528–580.

Cherry, J. R., D. Hondred, J. M. Walker, J. M. Keller, H. P. Hershey & R. D. Vierstra. 1993. Carboxy-terminal deletion of oat phytochrome A reveals the presence of separate domains required for structure and biological activity. Pl. Cell 5: 565–575.

Christensen, A. H. & P. H. Quail. 1989. Structure and expression of a maize phytochrome-encoding gene. Gene 85: 381–390.

Clack, T., S. Mathews & R. A. Sharrock. 1994. The phytochrome apoprotein family in *Arabidopsis* is encoded by five genes: The sequences and expression of *PHYD* and *PHYE*. Pl. Molec. Biol. 25: 413–427.

Cracraft, J. & K. Helm-Bychowski. 1991. Parsimony and methodological strategies. Pp. 184–220 *in* M. M. Miyamoto & J. Cracraft (editors), Phylogenetic Analysis of DNA Sequences. Oxford Univ. Press, New York.

Cronquist, A. 1981. An Integrated System of Classification of Flowering Plants. Columbia Univ. Press, New York.

Dehesh, K., J. Tepperman, A. H. Christensen & P. H. Quail. 1991. *phy*B is evolutionarily conserved and constitutively expressed in rice seedling shoots. Molec. Gen. Genet. 225: 305–313.

Doebley, J., M. Durbin, E. M. Golenberg, M. T. Clegg & D. P. Ma. 1990. Evolutionary analysis of the large subunit of carboxylase (*rbc*L) nucleotide sequence among the grasses (Gramineae). Evolution 44: 1097–1108.

Donoghue, M. J. 1994. Progress and prospects in reconstructing plant phylogeny. Ann. Missouri Bot. Gard. 81: 405–418.

Doyle, J. J. 1991. Evolution of higher-plant glutamine synthetase genes: Tissue specificity as a criterion for predicting orthology. Molec. Biol. Evol. 8: 366–377.

———. 1992. Gene trees and species trees: Molecular systematics as one-character taxonomy. Syst. Bot. 17: 144–163.

——— & J. L. Doyle. 1987. A rapid DNA isolation procedure for small amounts of fresh leaf tissue. Phytochem. Bull. 19: 11–15.

——— & J. L. Doyle. 1993. Chloroplast DNA phylogeny of the papilionoid legume tribe Phaseoleae. Syst. Bot. 18: 309–327.

Drouin, G. & G. A. Dover. 1990. Independent gene evolution in the potato actin gene family demonstrated by phylogenetic procedures for resolving gene conversions and the phylogeny of angiosperm actin genes. J. Molec. Evol. 31: 132–150.

Dunn, S. T. 1912. A revision of the genus *Millettia* Wight et Arn. J. Linn. Soc. (Bot.) 41: 123–243.

Edgerton, M. D. & A. M. Jones. 1992. Localization of protein-protein interactions between subunits of phytochrome. Pl. Cell 4: 161–171.

Evans, S. V., L. E. Fellows & E. A. Bell. 1985. Distribution and systematic significance of basic non-protein amino-acids and amines in the Tephrosieae (Leg.). Biochem. Syst. Ecol. 13: 271–302.

Farris, J. S. 1988. Hennig86. Computer program and documentation. Port Jefferson Station, New York.

Felsenstein, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39: 783–791.

———. 1993. PHYLIP—Phylogeny Inference Package, version 3.52c. Distributed by the author. Department of Genetics, Univ. Washington, Seattle.

Furuya, M. 1993. Phytochromes: Their molecular spe-

cies, gene families, and functions. Annual Rev. Pl. Physiol. Pl. Molec. Biol. 44: 617–645.

GAUT, B. S. & M. T. CLEGG. 1993. Nucleotide polymorphism in the *Adh1* locus of pearl millet (*Pennisetum glaucum*) (Poaceae). Genetics 135: 1091–1097.

GEESINK, R. 1981. Tephrosieae. Pp. 245–260 *in* R. M. Polhill & P. H. Raven (editors), Advances in Legume Systematics. Royal Botanic Gardens, Kew.

———. 1984. Scala Millettiearum. A survey of the genera of the tribe Millettieae (Legum.-Pap.) with methodological considerations. E. J. Brill/Leiden Univ. Press, Leiden.

GOODMAN, M., J. CZELUSNIAK, G. W. MOORE, A. E. ROMERO-HERRERA & G. MATSUDA. 1979. Fitting the gene lineage into its species lineage, a parsimony strategy illustrated by cladograms constructed from globin sequences. Syst. Zool. 28: 132–163.

HANELT, S., B. BRAUN, S. MARX & H. A. W. SCHNEIDER-POETSCH. 1992. Phytochrome evolution: A phylogenetic tree with the first complete sequence of phytochrome from a cryptogamic plant (*Selaginella martensii* Spring). Photochem. Photobiol. 56: 751–758.

HARRISON, R. G., D. M. RAND & W. C. WHEELER. 1987. Mitochondrial DNA variation in field crickets across a narrow hybrid zone. Molec. Biol. Evol. 4: 144–158.

HENDY, M. D. & D. PENNY. 1989. A framework for the quantitative study of evolutionary trees. Syst. Zool. 38: 297–309.

HERENDEEN, P. S. 1992. The fossil record of the Leguminosae from the Eocene of southeastern North America. Pp. 85–160 *in* P. S. Herendeen & D. L. Dilcher (editors), Advances in Legume Systematics, Part 4. Royal Botanic Gardens, Kew.

HERSHEY, H. P., R. F. BARKER, K. B. IDLER, J. L. LISSEMORE & P. H. QUAIL. 1985. Analysis of cloned cDNA and genomic sequences for phytochrome: Amino acid sequences for two gene products expressed in etiolated *Avena*. Nucl. Acids Res. 13: 8543–8559.

HEYER, A. & C. GATZ. 1992a. Isolation and characterization of a cDNA clone coding for potato type A phytochrome. Pl. Molec. Biol. 18: 535–544.

——— & ———. 1992b. Isolation and characterization of a cDNA clone coding for potato type B phytochrome. Pl. Molec. Biol. 20: 589–600.

IOERGER, T. R., A. G. CLARK & T. KAO. 1990. Polymorphism at the self-incompatibility locus in Solanaceae predates speciation. Proc. Natl. Acad. Sci. U.S.A. 87: 9732–9735.

IWABE, N., K. KUMA, M. HASEGAWA, S. OSAWA & T. MIYATA. 1989. Evolutionary relationship of archaebacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. Proc. Natl. Acad. Sci. U.S.A. 86: 9355–9359.

KAY, S. A., B. KEITH, K. SHINOZAKI & N.-H. CHUA. 1989. The sequence of the rice phytochrome gene. Nucl. Acids Res. 17: 2865–2866.

KIMURA, M. 1981. Estimation of evolutionary distances between homologous nucleotide sequences. Proc. Natl. Acad. Sci. U.S.A. 78: 454–458.

KLEIN, J., N. TAKAHATA & F. J. AYALA. 1993. MHC polymorphism and human origins. Sci. Amer. December 1993: 78–83.

KOLUKISAOGLU, H. U., B. BRAUN, W. F. MARTIN & H. A. W. SCHNEIDER-POETSCH. 1993. Mosses do express conventional, distantly B-type-related phytochromes. Phytochrome of *Physcomitrella patens* Hedw. Fed. Eur. Biochem. Soc. Lett. 334: 95–100.

KUMAR, S., K. TAMURA & M. NEI. 1993. MEGA: Molecular Evolutionary Genetics Analysis, version 1.01. The Pennsylvania State University, University Park, Pennsylvania.

LAVIN, M. & M. LUCKOW. 1993. Origins and relationships of tropical North America in the context of the boreotropics hypothesis. Amer. J. Bot. 80: 1–14.

LI, X. & M. NOLL. 1994. Evolution of distinct developmental functions of three *Drosophila* genes by acquisition of different *cis*-regulatory regions. Nature 367: 83–87.

LISTON, A. 1992. Use of the polymerase chain reaction to survey for the loss of the inverted repeat in the legume chloroplast genome. Third International Legume Conference, Poster Abstracts: 20. Royal Botanic Gardens, Kew.

MADDISON, W. P. & D. R. MADDISON. 1992. MacClade: Analysis of phylogeny and character evolution. Version 3.1. Sinauer Associates, Sunderland.

MAUCHER, H. P., R. SCHEUERLEIN & H. SCHRAUDOLF. 1992. Detection and partial sequence of phytochrome genes in the ferns *Anemia phyllitidis* (L.) Sw. (Schizaeaceae) and *Dryopteris filix-mas* L. (Polypodiaceae) by using the Polymerase-chain reaction technology. Photochem. Photobiol. 56: 759–763.

MCELROY, D., M. ROTHENBERG, K. S. REESE & R. WU. 1990. Characterization of the rice actin gene family. Pl. Molec. Biol. 15: 257–268.

MEAGHER, R. B., S. BERRY-LOWE & K. RICE. 1989. Molecular evolution of the small subunit of ribulose bisphosphate carboxylase: Nucleotide substitution and gene conversion. Genetics 123: 845–863.

MORAND, L. Z., D. G. KIDD & J. C. LAGARIAS. 1993. Phytochrome levels in the green alga *Mesotaenium caldariorum* are light-regulated. Pl. Physiol. 101: 97–103.

NAGATANI, A., J. W. REED & J. CHORY. 1993. Isolation and initial characterization of *Arabidopsis* mutants that are deficient in phytochrome A. Pl. Physiol. 102: 269–277.

NAGY, F., S. A. KAY & N.-H. CHUA. 1988. Gene regulation by phytochrome. Trends Genet. 4: 37–42.

NEI, M. & A. L. HUGHES. 1991. Polymorphism and evolution of the major histocompatibility complex loci in mammals. Pp. 222–247 *in* R. K. Selander, A. G. Clark & T. S. Whittam (editors), Evolution at the Molecular Level. Sinauer Associates, Sunderland, Massachusetts.

NGAI, J., M. M. DOWLING, L. BUCK, R. AXEL & A. CHESS. 1993. The family of genes encoding odorant receptors in the channel catfish. Cell 72: 657–666.

OKAMOTO, H., Y. HIRANO, H. ABE, K. TOMIZAWA, M. FURUYA & M. WADA. 1993. The deduced amino acid sequence of phytochrome from *Adiantum* includes consensus motifs present in phytochrome B from seed plants. Pl. Cell Physiol. 34: 1329–1334.

OLMSTEAD, R. G. 1989. Phylogeny, phenotypic evolution, and biogeography of the *Scutellaria angustifolia* complex (Lamiaceae): Inference from morphological and molecular data. Syst. Bot. 14: 320–338.

PAGE, R. D. M. 1990. Component analysis: A valiant failure? Cladistics 6: 119–136.

———. 1993. Component, version 2.0. Biogeography and Conservation Laboratory. The Natural History Museum, London.

PAMILO, P. & M. NEI. 1988. Relationships between gene trees and species trees. Molec. Biol. Evol. 5: 568–583.

PARKS, B. M. & P. H. QUAIL. 1993. *hy8*, a new class of *Arabidopsis* mutants deficient in functional phytochrome A. Pl. Cell 5: 39–48.

PATTERSON, C. 1987. Introduction. Pp. 1–22 *in* C. Patterson (editor), Molecules and Morphology in Evolution: Conflict or Compromise. Cambridge Univ. Press, Cambridge.

POLHILL, R. M. 1971. Some observation on generic limits in Dalbergieae-Lonchocarpinae. Kew Bull. 25: 259–273.

———. 1981. Dalbergieae. Pp. 233–242 *in* R. M. Polhill & P. H. Raven (editors), Advances in Legume Systematics, Part 1. Royal Botanic Gardens. Kew.

———. 1994. Classification of the Leguminosae. Chapter 1 *in* ILDIS Phytochemical Dictionary of the Leguminosae. Chapman Hall, London.

QUAIL, P. H. 1991. Phytochrome: A light-activated molecular switch that regulates plant gene expression. Annual Rev. Genet. 25: 389–409.

REED, J. W., P. NAGPAL, D. S. POOLE, M. FURUYA & J. CHORY. 1993. Mutations in the gene for the red/far-red light receptor phytochrome B alter cell elongation and physiological responses throughout *Arabidopsis* development. Pl. Cell 5: 147–157.

RIESEBERG, L. H. & S. J. BRUNSFELD. 1992. Molecular evidence and plant introgression. Pp. 151–176 *in* P. S. Soltis, D. E. Soltis & J. J. Doyle (editors), Molecular Systematics of Plants. Chapman and Hall, New York.

SANDERSON, M. J. & J. J. DOYLE. 1992. Reconstruction of organismal and gene phylogenies from data on multigene families: Concerted evolution, homoplasy and confidence. Syst. Biol. 41: 4–17.

SATO, N. 1988. Nucleotide sequence and expression of the phytochrome gene in *Pisum sativum*: Differential regulation by light of multiple transcripts. Pl. Molec. Biol. 11: 697–710.

———. 1990. Nucleotide sequence of a pseudogene for pea phytochrome reminiscent of an incorrect splicing event. Nucl. Acids Res. 18: 3632.

SHAH, D. M., R. C. HIGHTOWER & R. B. MEAGHER. 1983. Genes encoding actin in higher plants: Intron positions are highly conserved but the coding sequences are not. J. Molec. Appl. Genet. 2: 111–126.

SHARROCK, R. A. & P. H. QUAIL. 1989. Novel phytochrome sequences in *Arabidopsis thaliana*: Structure, evolution, and differential expression of a plant regulatory photoreceptor family. Genes and Development 3: 1745–1757.

———, J. L. LISSEMORE & P. H. QUAIL. 1986. Nucleotide and amino sequence of a *Cucurbita* phytochrome cDNA clone: Identification of conserved features by comparison with *Avena* phytochrome. Gene 47: 287–295.

SOLTIS, D. E., P. S. SOLTIS & B. G. MILLIGAN. 1992. Intraspecific chloroplast DNA variation: Systematic and phylogenetic implications. Pp. 117–150 *in* P. S. Soltis, D. E. Soltis & J. J. Doyle (editors), Molecular Systematics of Plants. Chapman and Hall, New York.

SOMERS, D. E., R. A. SHARROCK, J. M. TEPPERMAN & P. H. QUAIL. 1991. The *hy*3 long hypocotyl mutant of *Arabidopsis* is deficient in phytochrome B. Pl. Cell 3: 1263–1274.

SOUSA, M. & A. DELGADO. 1993. Mexican Leguminosae: Phytogeography, endemism, and origins. Pp. 459–511 *in* T. P. Ramamoorthy, R. Bye, A. Lot & J. Fa (editors), Biological Diversity of Mexico: Origins and Diversity. Oxford Univ. Press, Oxford.

——— & M. P. DE SOUSA. 1981. New World Lonchocarpinae. Pp. 261–281 *in* R. M. Polhill & P. H. Raven (editors), Advances in Legume Systematics, Part 1. Royal Botanic Gardens, Kew.

SWOFFORD, D. L. 1993. PAUP: Phylogenetic analysis using parsimony version 3.1. Illinois Natural History Survey, Champaign.

TAKAHATA, N. 1989. Gene genealogy in three related populations: Consistency probability between gene and population trees. Genetics 122: 957–966.

THOMAS, B. R., V. S. FORD, E. PICHERSKY & L. D. GOTTLIEB. 1993. Molecular characterization of duplicate cytosolic phosphoglucose isomerase genes in *Clarkia* and comparison to the single gene in *Arabidopsis*. Genetics 135: 895–905.

THÜMMLER, F., M. DUFNER, P. KREISL & P. DITTRICH. 1992. Molecular cloning of a novel phytochrome gene of the moss *Ceratodon purpureus* which encodes a putative light-regulated protein kinase. Pl. Molec. Biol. 20: 1003–1017.

TOWNROW, J. A. 1968. A fossil *Selaginella* from the Permian of New South Wales. J. Linn. Soc. (Bot.) 61: 13–23.

WAGNER, A., N. BLACKSTONE, P. CARTWRIGHT, M. DICK, B. MISOF, P. SNOW, G. P. WAGNER, J. BARTELS, M. MURTHA & J. PENDLETON. 1994. Surveys of gene families using polymerase chain reaction: PCR selection and drift. Syst. Biol. 43: 250–261.

WAYE, M. M. Y., M. E. VERHOEYEN, P. T. JONES & G. WINTER. 1985. *Eco*K selection vectors for shotgun cloning into M13 and deletion mutagenesis. Nucl. Acids Res. 13: 8561–8571.

WESTER, L., D. E. SOMERS, T. CLACK & R. A. SHARROCK. 1994. Transgenic complementation of the *hy3* phytochrome B mutation and response to *PHYB* gene copy number in *Arabidopsis*. Pl. J. 5: 261–272.

WHEELER, E. & P. BAAS. 1992. Fossil wood of the Leguminosae: A case study in xylem evolution and ecological anatomy. Pp. 281–301 *in* P. S. Herendeen & D. L. Dilcher (editors), Advances in Legume Systematics, Part 4. Royal Botanic Gardens, Kew.

WHITELAM, G. C., E. JOHNSON, J. PENG, P. CAROL, M. L. ANDERSON, J. S. COWL & N. P. HARBERD. 1993. Phytochrome A null mutants of *Arabidopsis* display a wild-type phenotype in white light. Pl. Cell 5: 757–768.

WINANDS, A., G. WAGNER, S. MARX & H. A. W. SCHNEIDER-POETSCH. 1992. Partial nucleotide sequence of phytochrome from the zygnematophycean green alga *Mougeotia*. Photochem. Photobiol. 56: 765–770.

WU, C.-I. & W.-H. LI. 1985. Evidence for higher rates

of nucleotide substitution in rodents than in man. Proc. Natl. Acad. Sci. U.S.A. 82: 1741–1745.

ZIMMER, E. A., S. L. MARTIN, S. M. BEVERLY, Y. W. KAN & A. C. WILSON. 1980. Rapid duplication and loss of genes coding for the alpha-chains of hemoglobin. Proc. Natl. Acad. Sci. U.S.A. 77: 2158–2162.

ZURAWSKI, G., M. T. CLEGG & A. H. D. BROWN. 1984. The nature of nucleotide sequence substitution between barley and maize chloroplast DNA. Genetics 106: 735–749.

Note added in proof.

Recent characterization of a genomic DNA clone from *Arabidopsis* containing *PHYC* indicates that this locus lacks the third intron shared by most of the other fully characterized *PHY* genes.

COWL, J. S., N. HARTLEY, D.-X. XIE, G. C. WHITELAM, G. P. MURPHY & N. P. HARBERD. 1994. The *PHYC* gene of *Arabidopsis*. Pl. Physiol. 106: 813–814.

```
                                  10         20         30         40         50
                                  *          *          *          *          *
Selaginella (sm) [1]              ---------- ---------- ---------- ---------- MSTTKLTYSS
Ceratodon (cp) [2]                ---------- ---------- ---------- ---------- MSATKKTYSS
Adiantum (ac) [3]                 ---------- ---------- ---------- ---------- MSSTRHSYSS
Arabidopsis PHYA (atA) [4]        ---------- ---------- ---------- -------MSG SRPT--QSSE
Cucurbita PHYA (cpA) [5]          ---------- ---------- ---------- -------MST SRPS--QSSS
Pisum PHYA (psA) [6]              ---------- ---------- ---------- -------MST TRPS--QSSN
Solanum PHYA (stA) [7]            ---------- ---------- ----MSTSLF ASDSDQLMSS SRPS--QSST
Avena PHYA (asA) [8]              ---------- ---------- ---------- -------MSS SRPA--SSSS
Oryza PHYA (osA) [9]              ---------- ---------- ---------- -------MSS SRPTQCSSSS
Zea PHYA (zmA) [10]               ---------- ---------- ---------- -------MSS SRPAHSSSSS
Arabidopsis PHYB (atB) [4]        --MVSGVGGS GGGRGGGRGG EEEPSSSHTP NNRRGGEQAQ SSGTKSLRPR
Arabidopsis PHYD (atD) [11]       MVSGGGSKTS GGEAASSGHR RSRHTSAAEQ AQSSANKALR WQNQQPQNHG
Arabidopsis PHYE (atE) [11]       ---------- ---------- ---------- ---------- MGFESSSSAA
Solanum PHYB (stB) [12]           ---------- ---------- -------MAS GSRTKHSHHS SSQAQSSGTS
Oryza PHYB (osB) [13]             MASGSRATPT RSPSSARPAA PRHQHHHSQS SGGSTSRAGG GGGGGGGGGG
Arabidopsis PHYC (atC) [4]        ---------- ---------- ---------- ---------- --MSSNTSRS
ANG                               ---------- ---------- ---------- ---------- ----------
CON                               ---------- ---------- ---------- ---------- ----------

                                 60         70         80         90         100        110        120
                                 *          *          *          *          *          *          *
sm   GSSAKSKHSV RVAQTTADAK LHAVYEESGE SGDSFDYSKS INATKSTGET IPAQ----AV -TAYLQRMQR
cp   TTSAKSKHSV RVAQTTADAA LEAVYEMSGD SGDSFDYSKS VGQSAE--SV P-----AGAV -TAYLQRMQR
ac   GGSGKSKHGR RIAQTSANAK LYAAYEESSE SGS-FDYSQS VSAGKEGI-- -----SSQLV -TAYLQRMQR
atA  GSRRSRHSAR IIAQTTVDAK LHADFE---E SGSSFDYSTS VRVTGPVV-- ENQPPRSDKV TTTYLHHIQK
cpA  NSGRSRHSTR IIAQTSVDAN VQADFE---E SGNSFDYSSS VRVTSDVS-- GDQQPRSDKV TTAYLHHIQK
psA  NSGRSRNSAR IIAQTTVDAK LHATFE---E SGSSFDYSSS VRVSGSVD-- GDQQPRSNKV TTAYLNHIQR
stA  TSSRSKHSAR IIAQTSIDAK LHADFE---E SGDSFDYSSS VRVTNVAE-- GEQRPKSDKV TTAYLHQIQK
asA  SRNRQSSQAR VLAQTTLDAE LNAEYE---E SGDSFDYSKL VEAQRDGP-- PVQQGRSEKV -IAYLQHIQK
osA  SRTRWSSRAR ILAQTTLDAE LNAEYE---E YGDSFDYSKL VEAQRTTG-- PEQQARSEKV -IAYLHHIQR
zmA  SRTRQSSRAR ILAQTTLDAE LNAEYE---E SGDSFDYSKL VEAQRSTP-- PEQQGRSGKV -IAYLQHIQR
atB  SNTESMSKSK AIQQYTVDAR LHAVFEQSGE SGKSFDYSQS LKTTTYGSSV PEQQ------ ITAYLSRIQR
atD  GGTESTNKNK AIQQYTVDAR LHAVFEQSGE SGKSFDYSQS LKTAPYDSSV PEQQ------ ITAYLSRIQR
atE  SNMKPQPQKS NTAQYSVDAA LFADFAQSIY TGKSFNYSKS VISPPN--HV PDEH------ ITAYLSNIQR
stB  NVNYKDSISK AIAQYTADAR LHAVFEQSGE SGKFFDYSQS VKTTTQ--SV PERQ------ ITAYLTKIQR
osB  GAAAAESVSK AVAQYTLDAR LHAVFEQSGA SGRSFDYTQS LRASPT--PS SEQQ------ IAAYLSRIQR
atC  CSTRSRQNSR VSSQVLVDAK LHGNFE---E SERLFDYSAS INLNM---PS SSCEIPSSAV -STYLQKIQR
ANG  ---------- ---Q---DA- ---------- ----F-Y--- ---------- ---------- ---YL--IQ-
CON  ---------- ---Q----A- ---------- ----F-Y--- ---------- ---------- ---YL---Q-

                                 130        140        150        160        170        180        190
                                 *          *          *          *          *          *          *
sm   GGLVQPFGCM LAV-EEGSFR VIAFSDNAGE MLDLMP-QSV PSL-GSGQQD VLTIGTDART LFTAAAS-AL
cp   EGLIQNFGCM VAV-EEPNFC VIAYSENASE FLDLIP-QAV PSM-GEM--D VLGIGTDIRT LFTPSSSAAL
ac   GGLVQQFGCL IAV-EEETFR VLHMCE-APE MLDVAT-QAV PTM-GQY--S RLCIGADVRT LLSPASASAL
atA  GKLIQPFGCL LAL-DEKTFK VIAYSENASE LLTMAS-HAV PSV-GEH--P VLGIGTDIRS LFTAPSASAL
cpA  GKLIQPFGCL LAL-DDKTFK VIAYSENAPE MLTMVS-HAV PSM-GDY--P VLGIGTDVRT IFTAPSASAL
psA  GKQIQPFGCL LAL-DEKTCK VVAYSENAPE MLTMVS-HAV PSV-GDH--P ALGIGTDIRT VFTAPSASAL
stA  GKFIQPFGCL LAL-DEKTLK VIAFSENAPE MLTMVS-HAV PSV-GEH--P VLGIGIDIRT IFTGPSGAAL
asA  GKLIQTFGCL LAL-DEKSFN VIAFSENAPE MLTTVS-HAV PSVD---DPP RLGIGTNVRS LFSDQGATAL
osA  AKLIQPFGCL LAL-DEKTFN VIALSENAPE MLTTVS-HAV PSVD---DPP KLRIGTNVRS LFTDPGTTAL
zmA  GKLIQPFGCL LAL-DEKSFR VIAFSENAPE MLTTVS-HAV PNVD---DPP KLGIGTNVRS LFTDPGATAL
atB  GGYIQPFGCM IAV-DESSFR IIGYSENARE MLGIMP-QSV PTLE---KPE ILAMGTDVRS LFTSSSSILL
atD  GGYTQPFGCL IAV-EESTFT IIGYSENARE MLGLMS-QSV PSIE-D-KSE VLTIGTDLRS LFKSSSYLLL
atE  GGLVQPFGCL IAV-EEPSFR ILGLSDNSSD FLGLLSLPST SHS-GEFDKV KGLIGIDART LFTPSSGASL
stB  GGHIQPFGCM IAV-DEASFR VIAYSENACE MLSLTP-QSV PSLE---KCE ILTIGTDVRT LFTPSSSVLL
osB  GGHIQPFGCT LAVADDSSFR LLAYSENTAD LLDLSPHHSV PSLDSSAVPP PVSLGADARL LFAPSSAVLL
atC  GMLIQPFGCL IVV-DEKNLK VIAFSENTQE MLGLIP-HTV PSME---QRE ALTIGTDVKS LFLSPGCSAL
ANG  ----Q-FGC- ---------- ----S-N--- -L-------- ---------- ----G----- -F-------L
CON  ----Q-FGC- ---------- ---------- -L-------- ---------- ----G----- ----------L
```

APPENDIX 1. All available full-length phytochrome amino acid sequences and 776 residues from *Ceratodon*. [1]Hanelt et al. (1992); [2]Thümmler et al. (1992); [3]Okamoto et al. (1993); [4]Sharrock & Quail (1989); [5]Sharrock et al. (1986); [6]Sato (1988); [7]Heyer & Gatz (1992a); [8]Hershey et al. (1985); [9]Kay et al. (1989); [10]Christensen & Quail (1989); [11]Clack et al. (1994); [12]Heyer & Gatz (1992b); [13]Dehesh et al. (1991). The triangle denotes the chromophore attachment site. The sequences amplified in this study correspond to residues 329–431.

```
                200        210        220        230        240        250        260
                 *          *          *          *          *          *          *
sm   EKAAGAVDLS MLNPIWVQSK TSAKPFYAIV HRIDVGLVMD LEPVKASDTR VGSAAGALQS HKLAAKAISR
cp   EKAAATQDIS LLNPITVHCR RSGKPLYAIA HRIDIGIVID FEAVKMIDVP VSAAAGALQS HKLAARAITR
ac   DRVIGVVDVS MFNPITVQSR SSGKPFYAIL HRNDVGLVID LEPIRPDDAS I-TG-GALQS HKLAAKAIAR
atA  QKALGFGDVS LLNPILVHCR TSAKPFYAII HRVTGSIIID FEPVKPYEVP M-TAAGALQS YKLAAKAITR
cpA  LKALGFGEVT LLNPILVHCK TSGKPFYAIV HRVTGSLIID FEPVKPYEGP V-TAAGALQS YKLAAKAITR
psA  QKALGFAEVS LLNPILVHCK TSGKPFYAII HRVTGSLIID FEPVKPYEVP M-TAAGALQS YKLAAKAITR
stA  QKALGFGEVS LLNPVLVHCK NSGKPFYAIV HRVTGSLIID FEPVKPYEVP M-TAAGALQS YKLAAKAITR
asA  HKALGFADVS LLNPILVQCK TSGKPFYAIV HRATGCLVVD FEPVKPTEFP A-TAAGALQS YKLAAKAISK
osA  QKALGFADVS LLNPILVQCK TSGKPFYAIV HRATGCLVVD FEPVKPTEFP A-TAAGALQS YKLAAKAISK
zmA  QKALGFADVS LLNPILVQCK TSGKPFYAIV HRATGCLVVD FEPVKPTEFP A-TAAGALQS YKLAAKAISK
atB  ERAFVAREIT LLNPVWIHSK NTGKPFYAIL HRIDVGVVID LEPAR-TEDP ALSIAGAVQS QKLAVRAISQ
atD  ERAFVAREIT LLNPIWIHSN NTGKPFYAIL HRVDVGILID LEPAR-TEDP ALSIAGAVQS QKLAVRAISH
atE  SKAASFTEIS LLNPVLVHSR TTQKPFYAIL HRIDAGIVMD LEPAK-SGDP ALTLAGAVQS QKLAVRAISR
stB  ERAFGAREIT LLNPIWIHSK NSGKPFYAIL HRVDVGIVID LEPAR-TEDP ALSIAGAVQS QKLRSEGLFL
osB  ERAFAAREIS LLNPLWIHSR VSSNPFYAIL HRIDVGVVID LEPAR-TEDP ALSIAGAVQS QKLVVRAISR
atC  EKAVDFGEIS ILNPITLHCR SSSKPFYAIL HRIEEGLVID LEPVSPDEVP V-TAAGALRS YKLAAKSISR
ANG  --A------- -LNP------ ----PFYAI- HR-------D -EP------- ----AGA--S -KL-------
CON  ---------- --NP------ ----P-YAI- HR-------D -E--------- -----GA--S -KL-------

                270        280        290        300        310        320        330
                 *          *          *          *          *          *          *
sm   LQSLP-GGDI GLLCDTVVEE VRDVTGYDLV MAYKFHEDEH GEVVAEIRRS DLEPYLGLHY PATDIPQASR
cp   LQALP-GGDI ELLCDTIVEE VRELTGYDRV MAFKFHEDEH GEVVAEIRRM DLEPYMGLHY PATDIPQASR
ac   LQSLP-GGDI GLLCDSVVEE VHELTGFDRV MAYKFHEDEH GEVVAEIRRT DLEPYIGLHY PATDIPQAAR
atA  LQSLP-SGSM ERLCDTMVQE VFELTGYDRV MAYKFHEDDH GEVVSEVTKP GLEPYLGLHY PATDIPQAAR
cpA  LQSLP-SGSM ARLCDTMVQE VFELTGYDRV MAYKFHDDDH GEVISEVAKP GLQPYLGLHY PATDIPQAAR
psA  LQSLA-SGSM ERLCDTMVQE VFELTGYDRV MAYKFHEDDH GEVIAEIAKP GLEPYLGLHY PATDIPQAAR
stA  LQSLP-SGSM ERLCDTMVQE VFELTGYDRV MGYKFHDDDH GEVVSEITKP GLEPYLGLHY PATDIPQAAR
asA  IQSLP-GGSM EVLCNTVVKE VFDLTGYDRV MAYKFHEDDH GEVFSEITKP GLEPYLGLHY PATDIPQAAR
osA  IQSLP-GGSM EVLCNTVVKE LFDLTGYDRV MAYKFHEDDH GEVFAEITKP GLEPYLGLHY PATDIPQAAR
zmA  IQSLP-GGSM EALCNTVVKE VFDLTGYDRV MAYKFHEDEH GEVFAEITKP GIEPYIGLHY PATDIPQAAR
atB  LQALP-GGDI KLLCDTVVES VRDLTGYDRV MVYKFHEDEH GEVVAESKRD DLEPYIGLHY PATDIPQASR
atD  LQSLP-SGDI KLLCDTVVES VRDLTGYDRV MVYKFHEDEH GEVVAESKRN DLEPYIGLHY PATDIPQASR
atE  LQSLP-GGDI GALCDTVVED VQRLTGYDRV MVYQFHEDDH GEVVSEIRRS DLEPYLGLHY PATDIPQAAR
stB  ICNHFLVGTL KLLCDTVVES VRELTGYDRV MVYKFHEDEH GEVVAESKRS DLEPYIGLHY PATDIPQASR
osB  LQALP-GGDV KLLCDTVVEH VRELTGYDRV MVYRFHEDEH GEVVAESRRS NLEPYIGLHY PATDIPQASR
atC  LQALP-SGNM LLLCDALVKE VSELTGYDRV MVYKFHEDGH GEVIAECCRE DMEPYLGLHY SATDIPQASR
ANG  --------G-- --LC---V-- ---LTGYDRV M-Y-FH-D-H GEV--E---- ---PY-GLHY -ATDIPQA-R
CON  --------G-- --LC---V-- ----TG-D-V M---FH-D-H GEV--E---- ---PY-GLHY -ATDIPQA-R
                                                                >TARGET
                340        350        360        370        380        390        400
                 *          *          *          *          ▼          *          *
sm   FLFMKNRVRM ICDCSAPPVK ITQDKELRQP ISLAGSTLRA PHGCHAQYMG NMGSVASLVM AMIINDNDE-
cp   FLLMKNRVRL IADCYASPVK LIQDPDIRQP VSLAGSTLRA PHGCHAQYMG NMGSIASLVM AVIINDNEE-
ac   FLFMKNRVRM ICDCRLPPVK LIQDKTLSQP MSLTGSTLRA PHGCHTQYMA NMNSISSLVM AVIVNDSDDD
atA  FLFMKNKVRM IVDCNAKHAR VLQDEKLSFD LTLCGSTLRA PHSCHLQYMA NMDSIASLVM AVVVNEEDGE
cpA  FLFMKNKVRM IVDCRAKHLK VLQDEKLQFD LTLCGSTLRA PHSCHLQYME NMNSIASLVM AVVVNEGDEE
psA  FLFMKNKVRM IVDCNAKHVK VLQDEKLPFD LTLCGSTLRA PHSCHLQYMA NMDSIASLVM AVVVNDSDED
stA  FLFMKNKVRM ICDCRAKHVK VVQDEKLPFD LTLCGSTLRA PHYCHLQYME NMNSIASLVM AVVVNDGDEE
asA  LLFMKNKVRM ICDCRARSIK VIEAEALPFD ISLCGSALRA PHSCHLQYME NMNSIASLVM AVVVNENEED
osA  FLFMKNKVRM ICDCRARSIK IIEDESLHLD ISLCGSTLRA PHSCHLQYME NMNSIASLVM AVVVNENEDD
zmA  FLFMKNKVRM ICDCRARSVK IIEDEALSID ISLCGSTLRA PHSCHLKYME NMNSIASLVM AVVVNENEED
atB  FLFKQNRVRM IVDCNATPVL VVQDDRLTQS MCLVGSTLRA PHGCHSQYMA NMGSIASLAM AVIINGNEDD
atD  FLFKQNRVRM IVDCYASPVR VVQDDRLTQF ICLVGSTLRA PHGCHAQYMT NMGSIASLAM AVIINGNEED
atE  FLFKQNRVRM ICDCNATPVK VVQSEELKRP LCLVNSTLRA PHGCHTQYMA NMGSVASLAL AIVVKGKD--
stB  FLFKQNRVRM IVDCHATPVR VTQDESLMQP LCLVGSTLRA PHGCHAQYMA NMGSIASLTL AVIINGNDEE
osB  FLFRQNRVRM IADCHAAPVR VIQDPALTQP LCLVGSTLRS PHGCHGQYMA NMGSIASLVM AVIISSGGDD
atC  FLFMRNKVRM ICDCSAVPVK VVQDKSLSQP ISLSGSTLRA PHGCHAQYMS NMGSVASLVM SVTINGSDSD
ANG  -LF--N-VRM I-DC-A---- ------L--- --L--S-LR- PH-CH--YM- NM-S-ASL-- ----------
CON  -L---N-VR- I-DC------ ---------- --L--S-LR- PH-CH--YM- NM-S--SL-- ----------
```

APPENDIX 1. Continued.

```
                410        420        430        440        450        460        470
                 *          *          *          *          *          *          *
sm   -PSGGGGGGG QHKGRRLWGL VVCHHTSPRS VPF-LRSACE FLMQVFGLQL NMEAAVAAHV REKHILRTQT
cp   -----YSRGA IQRGRKLWGL VVCQHTSPRT VPFPLRSVCE FLMQVFGMQL NLHVELAAQL REKHILRTQT
ac   ------SAGH SSQGIKLWGL VVCHHTSPRY VPFPVRSACE FLMQVFSLQL NMEVGMAAQV REKHILRTQT
atA GD-APDATTQ PQKRKRLWGL VVCHNTTPRF VPFPLRYACE FLAQVFAIHV NKEVELDNQM VEKNILRTQT
cpA NE---GPALQ QQKRKRLWGL VVCHNSSPRF VPFPLRYACE FLAQVFAIHV NKELELENQI IEKNILRTQT
psA GD--SADAVL PQKKKRLWGL VVCHNTTPRF VPFPLRYACE FLAQVFAIHV NKEIELEYQI LEKNILRTQT
stA GE--SSDSSQ SQKRKRLWGL VVSHNTTPRF APFPLRYACE FLAQVFAILV NKELELENQF LEKNILRTQT
asA DEAESEQPAQ QQKKKKLWGL LVCHHESPRY VPFPLRYACE FLAQVFAVHV NREFELEKQL REKNILKMQT
osA DEVGADQPAQ QQKRKKLWGL LVCHHESPRY VPFPLRYACE FLAQVFAVHV NKEFELERQV REKSILRMQT
zmA DEPEPEQPPQ QQKKKRLWGL IVCHHESPRY VPFPLRYACE FLAQVFAVHV NKEFELEKQI REKNILRMQT
atB G----SNVAS GRSSMRLWGL VVCHHTSSRC IPFPLRYACE FLMQAFGLQL NMELQLALQM SEKRVLRTQT
atD G---NGVNTG GRNSMRLWGL VVCHHTSARC IPFPLRYACE FFMQAFGLQL NMELQLASQV SEKRVLRMQT
atE ---------- ---SSKLWGL VVGHHCSPRY VPFPLRYACE FLMQAFGLQL QMELQLASQL AEKKAMRTQT
stB ------AVGG GRNSMRLWGL VVGHHTSVRS IPFPLRYACE FLMQAFGLQL NMELQLASQL SEKHVLRTQT
osB D--HNIARGS IPSAMKLWGL VVCHHTSPRC IPFPLRYACE FLMQAFGLQL NMELQLAHQL SEKHILRTGT
atC E-----MNRD LQTGRHLWGL VVCHHASPRF VPFPLRYACE FLTQVFGVQI NKEAESAVLL KEKRILQTQS
ANG ---------- ------LWGL -V-H----R- -PFPLRYACE F--Q-F---- --E------- -EK-------
CON ---------- ------LWGL -V------R- -PF--R--CE F--Q-F---- ---------- -EK-------
                                                                          <

                480        490        500        510        520        530        540
                 *          *          *          *          *          *          *
sm   LLCDMLLRDA -PIGIVSQSP NIMDLVKCDG AALYYGKRFW LLGITPSEAQ IKDIAEWLLE HH-KDSTGLS
cp   LLCDMLMRDA -PLGIVSQTP NIMDLVKCDG AALYYGKRVW LLGTTPTENQ IKEIADWLLE HH-MDSTGLS
ac   LLCDMLLRDA -PIGIVSQSP NIMDLVTCDG AALYYGKKCW LLGTTPTEAQ IVDIAAWLLD CH-KDSTGLS
atA LLCDMLMRDA -PLGIVSQSP NIMDLVKCDG AALLYKDKIW KLGTTPSEFH LQEIASWLCE YH-MDSTGLS
cpA LLCDMLMRDA -PLGIVSRSP NIMDLVKSDG AALLYKKKIW RLGLTPNDFQ LLDIASWLSE YH-MDSTGLS
psA LLCDMLMRDA -PLGIVSQSP NIMDLVKCDG AALFYRNKLW LLGATPTESQ LREIALWMSE YH-TDSTGLS
stA LLCDMLMRDA -PLGIVSQSP NIMDLIKCDG AALLYKNKIH RLGMNPSDFQ LHDIVSWLCE YH-TDSTGLS
asA MLSDMLFREA SPLTIVSGTP NIMDLVKCDG AALLYGGKVW RLQNAPTESQ IHDIAFWLSD VH-RDSTGLS
osA MLSDMLFRES SPLSIVSGTP NIMDLVKCDG AALLYGGKVW RLQNAPTESQ IRDIAFWLSD VH-RDSTGLS
zmA MLSDMLFKES SPLSIVSGSP NIMDLVKCDG AALLYGDKVW RLQTAPTESQ IRDIAFWLSE VH-GDSTGLS
atB LLCDMLLRDS -PAGIVTQSP SIMDLVKCDG AAFLYHGKYY PLGVAPSEVQ IKDVVEWLLA NH-ADSTGLS
atD LLCDMLLRDS -PAGIVTQRP SIMDLVKCNG AAFLYQGKYY PLGVTPTDSQ INDIVEWLVA NH-SDSTGLS
atE LLCDMLLRDT -VSAIVTQSP GIMDLVKCDG AALYYKGKCW LVGVTPNESQ VKDLVNWLVE NHGDDSTGLT
stB LLCDMLLRDS -PPGIVTQSP SIMDLVKCDG ALLYYQGKYY PLGVTPTEAQ IKDIVEWLLA YH-GDSTGLS
osB LLCDMLLRDS -PTGIVTQSP SIMDLVKCDG AALYYHGKYY PLGVTPTEVQ IKDIIEWLTM CH-GDSTGLS
atC VLCDMLFRNA -PIGIVTQSP NIMDLVKCDG AALYYRDNLW SLGVTPTETQ IRDLIDWVLK SH-GGNTGFT
ANG -L-DML---- ----IV---P -IMDL-K--G A---Y----- -----P---- ------W--- -H----TG--
CON -L-DML---- ----IV---P -IMDL----G A---Y----- -----P---- ------W--- -H----TG--

                550        560        570        580        590        600        610
                 *          *          *          *          *          *          *
sm   TDSLADAGYP GAASLGDEVC GMAAAKITAK DFLFWFRSHT AKEVKWGGAK HDPDDKDDGR KMHPRSSSKA
cp   TDSLADANYP GAHLLGDAVC GMAAAKITAK DFLFWFRSHT ATEVKWGGAK HDPDEKDDGR KMHPRSSFKA
ac   TDSLAKTGYP EASCLGDAVC GLAAAKITAT DFLFWFRSHT AKEVRWGGAR HDPEERDDGR RMHPRSSFKA
atA TDSLHDAGFP RALSLGDSVC GMAAVRISSK DMIFWFRSHT AGEVRWGGAK HDPDDRDDAR RMHPRSSFKA
cpA TDSLYDAGYP GAIALGDEVC GMAAVRITNN DMIFWFRSHT ASEIRWGGAK HEHGQKDDAR KMHPRSSFKA
psA TDSLSDAGFP GALSLSDTVC GMAAVRITSK DIVFWFRSHT AAEIRWGGAK HEPGDQDDGR KMHPRSSFKA
stA TDSLYDAGFP GALALGDAVC GMAAVRISDK DWLFWYRSHT AAEVRWGGAK HEPGEKDDGR KMHPRSSFKG
asA TDSLHDAGYP GAAALGDMIC GMAVAKINSK DILFWFRSHT AAEIRWGGAK NDPSDMDDSR RMHPRLSFKA
osA TDSLHDAGYP GAAALGDMIC GMAVAKINSK DILFWFRSHT AAEIRWGGAK HDPSDKDDSR RMHPRLSFKA
zmA TDSLQDAGYP GAASLGDMIC GMAVAKITSK DILFWFRSHT AAEIKWGGAK HDPSDKDDNR RMHPRLSFKA
atB TDSLGDAGYP GAAALGDAVC GMAVAYITKR DFLFWFRSHT AKEIKWGGAK HHPEDKDDGQ RMHPRSSFQA
atD TDSLGDAGYP RAAALGDAVC GMAVACITKR DFLFWFRSHT EKEIKWGGAK HHPEDKDDGQ RMNPRSSFQT
atE TDSLVDAGYP GAISLGDAVC GVAAAEFSSK DYLLWFRSNT ASAIKWGGAK HHPKDKDDAG RMHPRSSFTA
stB TDSLPDAGYP GAASLGDAVC GMAVAYITSK DFLFWFRSHT AKEIKWGGAK HHPEDKDDGQ RMHPRSSFKA
osB TDSLADAGYS GAAALGDAVS GMAVAYITPS DYLFWFRSHT AKEIKWGGAK HHPEDKDDGQ RMHPRSSFKA
atC TESLMESGYP DASVLGESIC GMAAVYISEK DFLFWFRSST AKQIKWGGAR HDPNDR-DGK RMHPRSSFKA
ANG T-SL---G-- -A--L----- G-A------- D---W-RS-T -----WGGA- --------D-- -M-PR-SF--
CON T-SL------ -A--L----- G-A------- D---W-RS-T -----WGGA- --------D-- -M-PR-S---
```

APPENDIX 1. Continued.

```
                620        630        640        650        660        670        680
                 *          *          *          *          *          *          *
sm   FLEVVKRRSL PWEDVEMDAI HSLQLILRGS FQDIDDSDTK TM-IHAR--- LNDLKLQGMD ELSTVANEMV
cp   FLEVVNKRSP PWEDVEMDAI HSLQLILRGS FRDIADSDTK TM-IHAR--- LNDLKLQGVE ERNALANEMS
ac   FLEVVKQQSL PWEDVEMDAI HSLQLILRGS FQDIDDSNTK TM-IHAR--- LNDLKLQGLD ELSTVASEMV
atA  FLEVVKTRSL PWKDYEMDAI HSLQLILRNA FKDSETTDVN TKVIYSK--- LNDLKIDGIQ ELEAVTSEMV
cpA  FLEVVKTRSL PWKDYEMDAI HSLQLILRNT FKDTDATEIN RKSIQTT--- LGDLKIEGRQ ELESVTSEMV
psA  FLEVVKARSV PWKDFEMDAI HSLQLILRNA SKDTDIIDLN TKAINTR--- LNDLKIEGMQ ELEAVTSEMV
stA  FLEVVKTRSI PWKDYEMDRI HSLQLILRNA FKDADAVNSN TISIHTK--- LNDLKIDGMQ ELEAVTAEMV
asA  FLEVVKMKSL PWSDYEMDAI HSLQLILRGT LNDASKPKRE ASLDNQI--- -GDLKLDGLA ELQAVTSEMV
osA  FLEVVKMKSL PWNDYEMDAI HSLQLILRGT LNDDIKPTRA ASLDNQV--- -GDLKLDGLA ELQAVTSEMV
zmA  FLEVVKTKSL PWSDYEMDAI HSLQLILRGT LNDASKPAQA SGLDNQI--- -GDLKLDGLA ELQAVTSEMV
atB  FLEVVKSRSQ PWETAEMDAI HSLQLILRDS FKESEAAMNS KVVDGVVQPC RDMAGEQGID ELGAVAREMV
atD  FLEVVKSRCQ PWETAEMDAI HSLQLILRDS FKESEAMDSK AAAAGAVQPH GDDMVQQGMQ EIGAVAREMV
atE  FLEVAKSRSL PWEISEIDAI HSLRLIMRES FTSSRPVLSG NGVARDAN-- ---------- ELTSFVCEMV
stB  FLEVVKSRSS PWENAEMDAI HSLQLILRDS FKDAEASNSK AIVHAH---- LGEMELQGID ELSSVAREMV
osB  FLEVVKSRSL PWENAEMDAI HSLQLILRDS FRDSAEGTSN SKAIVNGQVQ LGELELRGID ELSSVAREMV
atC  FMEIVRWKSV PWDDMEMDAI NSLQLIIKGS LQEEH---SK TVVDVP---- LVDNRVQKVD ELCVIVNEMV
ANG  F-E------- PW---E-D-I -SL-LI---- ---------- ---------- ---------- E------EMV
CON  F-E------- PW---E-D-I -SL-LI---- ---------- ---------- ---------- E------EM-

                690        700        710        720        730        740  INTRON 750
                 *          *          *          *          *          *          *
sm   RLIETATAPI LAVDSSGFIN GWNAKVADVT GLPVTEAMGR SLAKELVLHE SADMVERLLY LALQGDEEQN
cp   RVLETAAAPI LAVDSRGMIN AWNAKIAQVT GLPVEEAMHC SLTKDLVLDE SVVVVERLLS LALQGEEEQN
ac   RLIETATAPI LAVDGQGLIN GWNGKVAELT GLSFETAMGK SLAKELVHEE SKTIVERVLH LALEGEEEQD
atA  RLIETATVPI LAVDSDGLVN GWNTKIAELT GLSVDEAIGK HFLT-LVEDS SVEIVKRMLE NALEGTEEQN
cpA  RLIETATVPI LAVDLDGLIN GWNTKIAELT GLPVDKAIGK HLLT-LVEDS SVEVVRKMLF LALQGQEEQN
psA  RLIETATVPI LAVDVDGTVN GWNIKIAELT GLPVGEAIGK HLLT-LVEDS STDIVKKMLN LALQGEEEKN
stA  RLIETASVPI FAVDVDGQVN GWNTKVAELT GLPVDEAIGK HLLT-LVEDS SVDTVNKMLE LALQGQEERN
asA  RLMETATVPI LAVDGNGLVN GWNQKAAELT GLRVDDAIGR HILT-LVEDS SVPVVQRMLY LALQGKEEKE
osA  RLMETATVPI LAVDSNGLVN GWNQKVAELT GLRVDEAIGR HILT-VVEES SVPVVQRMLY LALQGKEEKE
zmA  RLMETATVPI LAVDGNGLVN GWNQKVAELS GLRVDEAIGR HILT-LVEDS SVSLVQRMLY LALQGREEKE
atB  RLIETATVPI FAVDAGGCIN GWNAKIAELT GLSVEEAMGK SLVSDLIYKE NEATVNKLLS RALRGDEEKN
atD  RLIETATVPI FAVDIDGCIN GWNAKIAELT GLSVEDAMGK SLVRELIYKE YKETVDRLLS CALKGDEGKN
atE  RVIETATAPI FGVDSSGCIN GWNKKTAEMT GLLASEAMGK SLADEIVQEE SRAALESLLC KALQGEEEKS
stB  RLIETATAPI FAVDVEGRIN GWNAKVAELT GVSVEEAMGK SLVHDLVYKE SQETAEKLLY NALRGEEDKN
osB  RLIETATVPI FAVDTDGCIN GWNAKVAELT GLSVEEAMGK SLVNDLIFKE SEETVNKLLS RALRGDEDKN
atC  RLIDTAAVPI FAVDASGVIN GWNSKAAEVT GLAVEQAIGK P-VSDLVEDD SVETVKNMLA LALEGSEERG
ANG  R---TA--PI --VD--G--N GWN-K-AE-- G-----A-G- ---------- ---------L- -AL-G-E---
CON  R---TA--PI --VD--G--N -WN-K-A--- G-----A--- ---------- ---------L- -AL-G-E---

                760        770        780        790        800        810        820
                 *          *          *          *          *          *          *
sm   VELKLKTFGG QKDKEAVIL- --VVNACASR DVSDNVVGVC FVGQDVTGQK VVMDKFTRIQ GDYKAIVQNP
cp   VEIKLKTFGT QTTERAVIL- --IVNACCSR DASDFVVGVF FVGQDVTEQR MFMDRFTRIQ GGEKTTVQDP
ac   IEIHLRTYDQ HKQKGVVIL- --IVNTCCSR DVSNNVVGVC FVGQDVTGQK LVLDRFIRIQ GDYKAIVQSL
atA  VQFEIKTHLS RADAGPISL- --VVNACASR DLHENVVGVC FVAHDLTGQK TVMDKFTRIE GDYKAIIQNP
cpA  VQFEIKTHGS HIEVGSISL- --VVNACASR DLRENVVGVF FVAQDITGQK MVMDKFTRLE GDYKAIVQNP
psA  VQFEIKTHGD QVESGPISL- --IVNACASK DLRENVVGVC FVAQDITAQK TVMDKFTRIE GDYKAIVQNP
stA  VEFEIKTHGP SRDSSPISL- --IVNACASK DVRDSVVGVC FIAQDITGQK SIMDKFTRIE GDYRAIIQNP
asA  VRFEVKTHGP KRDDGPVIL- --VVNACASR DLHDHVVGVC FVAQDMTVHK LVMDKFTRVE GDYKAIIHNP
osA  VKFEVKTHGS KRDDGPVIL- --VVNACASR DLHDHVVGVC FVAQDMTVHK LVMDKFTRVE GDYKAIIHNP
zmA  VRFELKTHGS KRDDGPVIL- --VVNACASR DLHDHVVGVC FVAQDMTVHK LVMDKFTRVE GDYKAIIHNP
atB  VEVKLKTFSP ELQGKAVFV- --VVNACSSK DYLNNIVGVC FVGQDVTSQK IVMDKFINIQ GDYKAIVHSP
atD  VEVKLKTFGS ELQGKAMFV- --VVNACSSK DYLNNIVGVC FVGQDVTGHK IVMDKFINIQ GDYKAIIHSP
atE  VMLKLRKFGQ NNHPDYSSDV CVLVNSCTSR DYTENIIGVC FVGQDITSEK AITDRFIRLQ GDYKTIVQSL
stB  VEIKLRTFGA EQLEKAVFV- --VVNACA-R DYTNNIVGVC FVGQDVTGEK VVMDKFINIQ GDYKAIVHSP
osB  VEIKLKTFGP EQSKGPIFV- --IVNACSTR DYTKNIVGVC FVGQDVTGQK VVMDKFINIQ GDYKAIVHNP
atC  AEIRIRAFGP KRKSSPVEL- --VVNTCCSR DMTNNVLGVC FIGQDVTGQK TLTENYSRVK GDYARIMWSP
ANG  ---------- ---------- ---VN-C--- D------GV- F---D-T--K ---------- GDY--I----
CON  ---------- ---------- ---VN-C--- D------GV- F---D-T--- ---------- G----------
```

APPENDIX 1. Continued.

```
          830        840        850        860        870        880        890
           *          *          *          *          *          *          *
sm  NPLIPPIFGA DEFGYCSEWN PAMEKLSGWR REEVLGKMLV GEIFGIQMMY CRLKGQDAVT KFMIVLNSAA
cp  HPLMRPSFDG DEFGRTFKRN SALGGL---- ---------- ---------- ---------- ----------
ac  NPLIPPIFGA DEYGFCSEWN AAMEKLSNWR REEVLGKMLV GEIFGLQMVC CRLQGQDVVT KLMIVLNDAV
atA NPLIPPIFGT DEFGWCTEWN PAMSKLTGLK REEVIDKMLL GEVFGTQKSC CRLKNQEAFV NLGIVLNNAV
cpA NPLIPPIFGS DEFGWCSEWN PAMAKLTGWS REEVIDKMLL GEVFGVHKSC CRLKNQEAFV NLGIVLNNAM
psA NQLIPPIFGT DEFGWCCEWN AAMIKLTGWK REEVMDKMLL GEVFGTQMSC CRLKNQEAFV NFGIVLNKAM
stA HPLIPPIFGT DQFGWCSEWN SAMTMLTGWR RDDVMDKMLL GEVFGTQAAC CRLKNQEAFV NFGVILNNAI
asA NPLIPPIFGA DEFGWCSEWN AAMTKLTGWN RDEVLDKMLL GEVFDSSNAS CPLKNRDAFV SLCVLINSAL
osA SPLIPPIFGA DEFGWCSEWN AAMTKLTGWH RDEVINKMLL GEVFDSTNAS CLVKNKDAFV SLCILINSAL
zmA NPLIPPIFGA DQFGWCSEWN AAMTKLTGWH RDEVVDKMLL GEVFNSSNAS CLLKSKDAFV RLCIVINSAL
atB NPLIPPIFAA DENTCCLEWN MAMEKLTGWS RSEVIGKMIV GEVFG---SC CMLKGPDALT KFMIVLHNAI
atD NPLIPPIFAA DENTCCLEWN TAMEKLTGWP RSEVIGKLLV REVFG---SY CRLKGPDALT KFMIVLHNAI
atE NPLIPPIFAS DENACCSEWN AAMEKLTGWS KHEVIGKMLP GEVFG---VF CKVKCQDSLT KFLISLYQGI
stB NPLIPPIFAS DENTCCSEWN TAMEKLTGWS RGEIVGKMLV GEIFG---SC CRLKGPDAMT KFMIVLHNAI
osB NPLIPPIFAS DENTCCSEWN TAMEKLTGWS RGEVVGKLLV GEVFG---NC CRLKGPDALT KFMIVLHNAI
atC STLIPPIFIT NENGVCSEWN NAMQKLSGIK REEVVNKILL GEVFTTDDYG CCLKDHDTLT KLRIGFNAVI
ANG --LIPPIF-- -----C-EWN -AM-KL-G-- ------K--- -E-F------ C--K------ ----------
CON --L--P-F-- ---------N -A---L---- ------K--- -E-F------ C--------- ----------

          900        910        920        930        940        950        960
           *          *          *          *          *          *          *
sm  DGQ-DTEKFP FAFFDRQGKY VEALLTATKR ADAEGSITGV FCFPHIASAE LQQALTVQRA TEKVALSKLK
cp  ---------- ---------- ---------- ---------- ---------- ---------- ----------
ac  NGQ-ESEKFP LVFYDRNGRR VEALLIASKR TDADGRITGV FCFLHTASPE LLQALIIKRA KEKV----DK
atA TSQ-DPEKVS FAFFTRGGKY VECLLCVSKK LDREGVVTGV FCFLQLASHE LQQALHVQRL AERTAVKRLK
cpA CGQ-DPEKAS FGFLARNGMY VECLLCVNKI LDKDGAVTGF FCFLQLPSHE LQQALNIQRL CEQTALKRLR
psA TGL-ETEKVP FGFFSRKGKY VECLLSVSKK IDAEGLVTGV FCFLQLASPE LQQALHIQRL SEQTALKRLK
stA TGQ-ESEKIP FGFFARYGKY VECLLCVSKR LDKEGAVTGL FCFLQLASHE LQQALHVQRL SEQTALKRLK
asA AGE-ETEKAP FGFFDRSGKY IECLLSANRK ENEGGLITGV FCFIHVASHE LQHALQVQQA SEQTSLKRLK
osA AGD-ETEKAP FSFFDRNGKY IECLLSVNRK VNADGVITGV FCFIQVPSHE LQHALHVQQA SQQNALTKLK
zmA AGE-EAEKAS FGFFDRNEKY VECLLSVNRK VNADGVVTGV FCFIHVPSDD LQHALHVQQA SEQTAQRKLK
atB GGQ-DTDKFP FPFFDRNGKF VQALLTANKR VSLEGKVIGA FCFLQIPSPE LQQALAVQRR QDTECFTKAK
atD GGQ-DTDKFP FPFFDRKGEF IQALLTLNKR VSIDGKIIGA FCFLQIPSPE LQQALEVQRR QESEYFSRRK
atE AGDNVPESSL VEFFNKEGKY IEASLTANKS TNIEGKVIRC FFFLQIINKE SGLSCPELKE SAQS----LN
stB GGQ-DTDKFP FSFFDRNGKY VQALLTRNKR VNMEGDTIGA FCFIQIASPE LQQALRVQRQ QEKKCYSQMK
osB GGQ-DCEKFP FSFFDKNGKY VQALLTANTR SRMDGEAIGA FCFLQIASPE LQQAFEIQRH HEKKCYARMK
atC SGQKNIEKLL FGFYHRDGSF IEALLSANKR TDIEGKVTGV LCFLQVPSPE LQYALQVQQI SEHAIACALN
ANG ---------- --F------- ----L----- ----G----- --F------- ---------- ----------
CON ---------- --F------- ----L----- ----G----- --F------- ---------- ----------

          970        980        990       1000       1010       1020 INTRON 1030
           *          *          *          *          *          *    |     *
sm  ELAYIRQEIK NPLYGIMFTR TLMETTDLSK DQKQYFETGA VCEKQIRKIL DDMDLESIED G--YLELDTT
cp  ---------- ---------- ---------- ---------- ---------- ---------- ----------
ac  ELSYVKEELK KPLEGLAFTR TVLEGTNLTI EQRQLIKTNA WCERQLRKIL -EDDLNNIEE G--YMDLEMS
atA ALAYIKRQIR NPLSGIMFTR KMIEGTELGP EQRRILQTSA LCQKQLSKIL DDSDLESIIE G--CLDLEMK
cpA ALGYIKRQIQ NPLSGIIFSR RLLERTELGV EQKELLRTSG LCQKQISKVL DESDIDKIID G--FIDLEMD
psA VLTYMKRQIR NPLAGIVFSS KMLEGTDLET EQKRIVNTSS QCQRQLSKIL DDSDLDGIID G--YLDLEMA
stA VLAYIRRQIR NPLSGIIFSR KMLEGTSLGE EQKNILHTSA QCQRQLDKIL DDTDLDSIIE G--YLDLEML
asA AFSYMRHAIN NPLSGMLYSR KALKNTDLNE EQMKQIHVGD NCHHQINKIL ADLDQDSITE KSSCLDLEMA
osA AYSYMRHAIN NPLSGMLYSR KALKNTGLNE EQMKEVNVAD SCHRQLNKIL SDLDQDSVMN KSSCLDLEMV
zmA AFSYMRHAIN KPLSGMLYSR ETLKSTGLNE EQMRQVRVGD NCHRQLNKIL ADLDQDNITD KSSCLDLDMA
atB ELAYICQVIK NPLSGMRFAN SLLEATDLNE DQKQLLETSV SCEKQISRIV GDMDLESIED G--SFVLKRE
atD ELAYIFQVIK NPLSGLRFTN SLLEDMDLNE DQKQLLETSV SCEKQISKIV GDMDVKSIDD G--SFLLERT
atE ELTYVRQEIK NPLNGIRFAH KLLESSEISA SQRQFLETSD ACEKQITTII ESTDLKSIEE G--KLQLETE
stB ELAYICQEIK SPLNGIRFTN SLLEATNLTE NQKQYLETSA ACERQMSKII RDIDLENIED G--SLTLEKE
osB ELAYIYQEIK NPLNGIRFTN SLLEMTDLKD DQRQFLETST ACEKQMSKIV KDASLQSIED G--SLVLEKG
atC KLAYLRHEVK DPEKAISFLQ DLLHSSGLSE DQKRLLRTSV LCREQLAKVI SDSDIEGIEE G--YVELDCS
ANG ---Y------ -P-------- ---------- -Q-------- -C--Q----- ---------- ------L---
CON ---Y------ -P-------- ---------- -Q-------- -C--Q----- ---------- ------L---
```

APPENDIX 1. Continued.

```
          1040        1050        1060        1070        1080        1090        1100
             *           *           *           *           *           *           *
sm   EFMMGTVMDA VISQGMITSK EKNLQLIRET PKEIKAMFLY GDQVRLQQVL ADFLLNAIRF TPSSEN----
cp   ---------- ---------- ---------- ---------- ---------- ---------- ----------
ac   EFFMGSVIDA VISQGMAASR GKGVQILTEI PNDVKLMCLF GDQARLQQVL ADLLFCAINH ATTTNEDEKD
atA  EFTLNEVLTA STSQVMMKSN GKSVRITNET GEEVMSDTLY GDSIRLQQVL ADFMLMAVNF TPSGG-----
cpA  EFTLHEVLMV SISQVMLKIK GKGIQIVNET PEEAMSETLY GDSLRLQQVL ADFLLISVSY APSGG-----
psA  EFTLHEVLVT SLSQVMNRSN TKGIRIANDV AEHIARETLY GDSLRLQQVL ADFLLISINS TPNGG-----
stA  EFKLHEVLVA SISQVMMKSN GKNIMISNDM VEDLLNETLY GDSPRLQQVL ANFLLVSVNS TPSGG-----
asA  EFLLQDVVVA AVSQVLITCQ GKGIRISCNL PERFMKQSVY GDGVRLQQIL SDFLFISVKF SPVGG-----
osA  EFVLQDVFVA AVSQVLITCQ GKGIRVSCNL PERYMKQTVY GDGVRLQQIL SDFLFVSVKF SPVGG-----
zmA  EFVLQDVVVS AVSQVLIGCQ AKGIRVACNL PERSMKQKVY GDGIRLQQIV SDFLFVSVKF SPAGG-----
atB  EFFLGSVINA IVSQAMFLLR DRGLQLIRDI PEEIKSIEVF GDQIRIQQLL AEFLLSIIRY APSQE-----
atD  EFFIGNVTNA VVSQVMLVVR ERNLQLIRNI PTEVKSMAVY GDQIRLQQVL AEFLLSIVRY APMEG-----
atE  EFRLENILDT IISQVMIILR ERNSQLRVEV AEEIKTLPLN GDRVKLQLIL ADLLRNIVNH APFPNS----
stB  DFFLGSVIDA VVSQVMLLLR EKGVQLIRDI PEEIKTLTVH GDQVRIQQVL ADFLLNMVRY APSPDG----
osB  EFSLGSVMNA VVSQVMIQLR ERDLQLIRDI PDEIKEASAY GDQYRIQQVL CDFLLSMVRF APAENG----
atC  EFGLQESLEA VVKQVMELSI ERKVQISCDY PQEVSSMRLY GDNLRLQQIL SETLLSSIRF TPALRGL---
ANG  -F--------- ---Q------ ---------- ---------- GD----Q--- ---------- -P--------
CON  -F--------- ---Q------ ---------- ---------- GD----Q--- ---------- ----------

          1110        1120  INTRON    1130        1140        1150        1160        1170
             *           *     |     *           *           *           *           *
sm   WVGIKVATSR KRLGGVVHVM HLEFRITHPG VGLPEELVQE MFDRGRGM-T QEGLGLSMCR KLVKLMN-GE
cp   ---------- ---------- ---------- ---------- ---------- ---------- ----------
ac   WVTIKVSRTK TRLDDGVHLM HFESRISHSG QGISEALVEE MTNKSQKW-T PEGLAISISC TLIRLMN-GD
atA  QLTVSASLRK DQLGRSVHLA NLEIRLTHTG AGIPEFLLNQ MFGTEE-DVS EEGLSLMVSR KLVKLMN-GD
cpA  QLTISTDVTK NQLGKSVHLV HLEFRITYAG GGIPESLLNE MFGSEE-DAS EEGFSLLISR KLVKLMN-GD
psA  QVVIAASLTK EQLGKSVHLV NLELSITHGG SGVPEAALNQ MFGNNV-LES EEGISLHISR KLLKLMN-GD
stA  KLSISGKLTK DRIGESVQLA LLEFRIRHTG GGVPEELLSQ MFGSEA-DAS EEGISLLVSR KLVKLMN-GE
asA  SVEISSKLTK NSIGENLHLI DLELRIKHQG LGVPAELMAQ MFEEDNKEQS EEGLSLLVSR NLLRLMN-GD
osA  SVEISCSLTK NSIGENLHLI DLELRIKHQG KGVPADLLSQ MYEDDNKEQS DEGMSLAVSR NLLRLMN-GD
zmA  SVDISSKLTK NSIGENLHLI DFELRIKHRG AGVPAEILSQ MYEEDNKEQS EEGFSLAVSR NLLRLMN-GD
atB  WVEIHLSQLS KQMADGFAAI RTEFRMACPG EGLPPELVRD MFHSSR-WTS PEGLGLSVCR KILKLMN-GE
atD  SVELHLCPTL NQMADGFSAV RLEFRMACAG EGVPPEKVQD MFHSSR-WTS PEGLGLSVCR KILKLMN-GG
atE  WVGISISPGQ ELSRDNGSRI HLQFRMIHPG KGLPSEMLSD MFETRDGWVT PDGLGLKLSR KLLEQMN-GR
stB  WVEIQLRPSM MPISDGVTVV HIELGLYAPG -RLPPELVQD MFHSSR-WVT QEGLGLSMCR KMLKLMN-GE
osB  WVEIQVRPNI KQNSDGTDTM LFPFRFACPG EGLPPEIVQD MFSNSR-WTT QEGIGLSICR KILKLMG-GE
atC  CVSFKVIARI EAIGKRMKRV ELEFRIIHPA PGLPEDLVRE MFQPLRKGTS REGLGLHITQ KLVKLMERGT
ANG  ---------- ---------- ---------- ---P------ M--------- --G--L---- -----M----
CON  ---------- ---------- ---------- ---------- M--------- --G------- -----M----

          1180        1190        1200        1210
             *           *           *           *
sm   VEYIREAGKN YFLVSLELPL AQRDDAGSVK FQASS----- -
cp   ---------- ---------- ---------- ---------- -
ac   VKYTTDAGNK CFLVTIQFPL AHRDDATSVR ---------- -
atA  VQYLRQAGKS SFIITAELAA ANK------- ---------- -
cpA  VRYMREAGKS SFIITVELAA AHKSRTT--- ---------- -
psA  VRYLKEAGKS SFILSVELAA AHKLKG---- ---------- -
stA  VQYLREAGRS TFIISVELAV ATKSS----- ---------- -
asA  VRHLREAGVS TFIITAELAS APTAMGQ--- ---------- -
osA  VRHMREAGMS TFILSVELAS APAK------ ---------- -
zmA  IRHLREAGMS TFILTAELAA APSAVGR--- ---------- -
atB  VQYIRESERS YFLIILELPV PRKRPLSTAS GSGDMMLMMP Y
atD  VQYIREFERS YFLIVIELPV PLMMMMPSS- ---------- -
atE  VSYVREDERC FFQVDLQVKT MLGVESRGTE GSSSIK---- -
stB  IQYIRESERC YFLIILDLPM TRKGPKSVG- ---------- -
osB  VQYIRESERS FFHIVLELPQ PQQAASRGTS ---------- -
atC  LRYLRESEMS AFVILTEFPL I--------- ---------- -
ANG  ---------- -F-------- ---------- ---------- -
CON  ---------- -F-------- ---------- ----------
```

APPENDIX 1. Continued.

```
                 300        310        320        330        340        350        360
                   *          *          *          *          *          *          *
Mougeotia[1]     DEHGEVVAEIRRSDLEPYLGLHYPATDIPQASRFLFIKNRIRMICDCTSPQVKVVQDSRIPQEMS
Ceratodon[2]     DEHGEVVAEIRRMDLEPYMGLHYPATDIPQASRFLLMKNRVRLIADCYASPVKLIQDPDIRQPVS
Selaginella[3]   DEHGEVVAEIRRSDLEPYLGLHYPATDIPQASRFLFMKNRVRMICDCSAPPVKITQDKELRQPIS
Psilotum[3]      DEHGEVVAEIRRSDLEPFVGIHYPATDIPQACRFLFLKNRVTMICDCYAPPIRIIQDRQLKQPLS
Adiantum[4]      DEHGEVVAEIRRTDLEPYIGLGYPATDIPQAARFLFMKNRVRMICDCRLPPVKLIQDKTLSQPMS
Anemia[5]        DEHGEVVAEIRRSDLEPYMGLHYPATDIPQAARFLFMKNRVRLIYDCRLPPVKVIQDKNLVQPLS
Dryopteris[5]    DEHGEVLAEIRRSDLEPYLGLHYPATDIPQASRFLFMKNRVRMICDCRAIPVRVIQDKELRQPLS

                 370        380        390        400        410        420
                   *          *          *          *          *          *
Mougeotia        LAGSTMRGVHGCHTQYMMNMGSTASLVMCVTINDTNE------IAGGPGMKGRKLWGLIVCHHST
Ceratodon        LAGSTLRAPHGCHAQYMGNMGSIASLVMAVIINDNEE------YSRGAIQRGRKLWGLVVCQHTS
Selaginella      LAGSTLRAPHGCHAQYMGNMGSVASLVMAMIINDNDE--PSGGGGGGGQHKGRRLWGLVVCHHTS
Psilotum         LAGSTLRAPHGCHAHYMGNMGSIASLVMAVIVKRHGEED---RSLGFQSQNGNRLWGMVVCHHTT
Adiantum         LTGSTLRAPHGCHTQYMANMNSISSLVMAVIVNDSDDD------SAGHSSQGIKLWGLVVCHHTS
Anemia           LAGSTLRAPHRCHAEYMGNMGSIASLGMAVIVNDDDSSD------AGNMQQRTRLWGLVVCHHTS
Dryopteris       LAGSTLRAPHGCHGQYMANMGSIASLVMAVVVNDNDED----LSNRPHQPKMRRLWGLVVCHHTT

                 430        440        450        460        470        480        490
                   *          *          *          *          *          *          *
Mougeotia        PRHIPFPIHSACEFLMQVFGLQLNMEAELAAQHREKHILRTQTLLCDMLLRDA-PMGIVSQSPN
Ceratodon        PRTVPFPLRSVCEFLMQVFGMQLNLHVELAAQLREKHILRTQTLLCDMLLRDA-PIGIVSQTPN
Selaginella      PRSVPF-LRSACEFLMQVFGLQLNMEAAVAAHVREKHILRTQTLLCDMLLRDA-PIGIVSQSPN
Psilotum         PRAVPFALRCACEFFAQVFALQLNMELELAAQMREKDILRTQSLLCDMLLRDA-PIGIVTRSPN
Adiantum         PRYVPFPVRSACEFLMQVFSLQLNMEVGMAAQVREKHILRTQTLLCDMLLRDA-PIGIVSQSPN
Anemia           TRYVPFPLRSACEFLMQVFSLELNMEVELAAQRREKHILQTQTLLCDMLLRDA-PIGIVSQSPN
Dryopteris       PRAVPFALRSACEFLMQVFGLQINMELELAAQMREKHILRTQTLLCDMLLRDA-PIGIVSESPN
```

APPENDIX 2. Regions of all nonangiosperm amino acid sequences available that are homologous with the *Mougeotia* gene fragment, numbered with reference to Appendix 1. [1]Winands et al. (1992); [2]Thümmler et al. (1992); [3]Hanelt et al. (1992); [4]Okamoto et al. (1993); [5]Maucher et al. (1992).

APPENDIX 3. Sources of *PHY* sequences determined in this study. Arrangement of flowering plants follows Cronquist (1981) and Polhill (1994).

Subclass/Tribe	Species	Source/Voucher
Sphenophyta	*Equisetum arvense* L.	P. Soltis (no voucher)
Pinophyta	*Gingko biloba* L.	*S. Mathews 365* MONT
	Pseudotsuga menziesii (Mirb.) Franco	*S. Mathews s.n.* MONT
Magnoliophyta		
Monocots		
Alismatidae	*Elodea* Michx. sp.	S. Mathews (no voucher)
Arecidae	*Lemna gibba* L.	J. Silverthorne (no voucher)
Commelinidae	*Hordeum vulgare* L.	*S. Mathews s.n.* MONT
	Calamovilfa longifolia (Hook.) Scribn.	*Lavin s.n.* MONT
	Panicum capillare L.	*Lavin s.n.* MONT
Zingiberidae	*Billbergia nutans* H. Wendl	*S. Mathews 351* MONT
Liliidae	*Muscari* Mill. sp.	S. Mathews (no voucher)
Dicots		
Magnoliidae	*Ceratophyllum demersum* L.	*S. Mathews s.n.* MONT
	Aquilegia L. sp.	S. Mathews (no voucher)
Hamamelidae	*Urtica dioca* L.	*S. Mathews 330* MONT
	Quercus turbinella Greene	*J. M. Tucker 4491* UCD
Caryophyllidae	*Dianthus caryophyllus* L.	R. Woodson (no voucher)
	Spinacia oleracea L.	S. Mathews (no voucher)
Dilleniidae	*Arabidopsis thaliana* (L.) Schur	S. Mathews (no voucher)
Asteridae	*Lycopersicon esculentum* Mill.	S. Mathews (no voucher)
	Antirrhinum majus L.	*S. Mathews 301* MONT
Rosidae	*Daucus carota* L.	S. Mathews (no voucher)
Fabaceae		
Dalbergieae	*Dalbergia* L.f. sp.	*Lavin 7141* MONT
	Tipuana tipu (Benth.) Kuntze	*Lavin 6184* BH
Galegeae	*Caragana arborescens* Lam.	*Lavin 5907* RM
	Clianthus formosus (G. Don) Ford & Vick	Krukoff s.n. K
Millettieae	*Dalbergiella nyasae* Baker f.	*Muller 2686* K
	Derris elliptica (Wallich) Benth.	Michigan State Univ. Conservatory (no voucher)
	Kunstleria blackii (F. Muell.) Prain	*Pedley 5005* K
	Lonchocarpus eriocarinalis Micheli	*Lavin 5325a* BH
	Millettia dura Dunn	*Lock 83/124* K
	Millettia richardiana (Baill.) D. J. Du Puy & J. Labat	*Schrire et al. 2555* K
	Piscidia piscipula (L.) Sarg.	*Lavin & Luckow 5793a* TEX
	Wisteria floribunda (Willd.) DC	*Lavin 6205* BH
	Xeroderris stuhlmanii (Taub.) Mendonca & E. P. Sousa	*Corby 2162* K
Robinieae	*Hebestigma cubense* (HBK) Urb.	*Lavin 5611* TEX
	Lennea melanocarpa (Schltdl.) Vatke ex Harms	*Lavin & Delgado 8217* MEXU
	Sesbania sesban (L.) Merr.	*Potter 870410* BH
	Sesbania vesicaria (Jacq.) Elliot	*Lavin s.n.* TEX
Sophoreae	*Myrospermum sousanum* A. Delgado & M. C. Johnston	*Delgado & Johnston s.n.* TEX
Vicieae	*Lathyrus odoratus* L.	*Lavin 6170* MONT

```
                          330       340       350       360       370       380       390       400                 410
                           *         *         *         *         *         *         *         *                   *
Myrospermum B   QASRFLFKQNRVRMIVDCNATPVRVIQDERLRQPLCLVGSTLRAPHGCHAQYMQKMGSIASLVMAVIINGND--EEAV-GG-----RSSMKLWGLVVCHH
Xeroderris B    QASRFLFKQNRVRMIVDCHASPVSVVQDEALVQPLCLVGSTLRAPHGCHAQYMANMGSIASLVMAVIINGND--EEGV-GG-----RSSMRLWGLVVCHH

Myrospermum A   QAARFLFMKNKVRMIVDCHAKHVKVLQDEKLPFDLTLCGSTLRAPHSCHLQYMANMDSIASLVLAVVVNDSD--EDGD-SSDAVQPQKRKRLWGLVVCHN
Poitea A        QAARFLFMKNKVRMIVDCHAKHVKVLQDEKLAFDLTLCGSTLRAPHSCHLQYMANMDSIASLVMAVVVNDSD--EDGD-SSDAIQPQKRKRLWGLVVCHN
Hybosema A      QAARFLFMKNKVRMIVDCHAKHVKVLQDEKLPFDLTLCGSTLRAPHSCHLQYMANMDSIASLVMAVVVNDSD--EDGD-SSDAVQPQKRKRLWGLVVCHN
Hebestigma A    QAARFLFMKNKVRMIVDCHAKHVKVLQDEKLPFDLTLCGSTLRAPHSCHLQYMANMDSIASLVMAVVVNDSD--EDGD-SSDAVQPQKRKRLWGLVVCHN
Lennea A        QAARFLFMKNKVRMIVDCHAKHVKVLQDEKLPFDLTLCGSTLRAPHSCHLQYMANMDSIASLVMAVVVNDSD--EDGD-SSDAVQPQKRKRLWGLVVCHN
Sesbanias A     QAARFLFMKNKVRMIVDCHAKHVKVLQDEKLPFDLTLCGSTLRAPHSCHLQYMANMDSIASLVMAVVVNDSD--EDGD-SSDAVQPQKRKRLWGLVVCHN
Sesbaniav A     QAARFLFMKNKVRMIVDCHAKHVKVLQDEKLPFDLTLCGSTLRAPHSCHLQYMANMDSIASLVMAVVVNDSD--EDGD-SSDAVQPQKRKRLWGLVVCHN
Piscidia A      QAARFLFMKNKVRMIVDCHAKHVKVLQDEKLPFDLTLCGSTLRAPHSCHLQYMANMDSIASLVMAVVVNDNE--ENGD-GSDAVQPQKRKRLWGLVVCHN
Kunsteria A     QAARFLFMKNKVRMIVDCHAKHVKVLQDEKLPFDLTLCGSTLRAPHSCHLQYMANMNSIASLVLAVVVNDNE--EDGD-SSDAVQPQKRKRLWGLVVCHN
Dalbergiella A  QAARFLFMKNKVRMIVDCHAKHVKVLQDEKLPLDLTLCGSTLRAPHSCHLQYMANMDSIASLVMAVVVNDNE--EDGD-SSDAVQPQKRKRLWGLVVCHN
Dalbergia A     QASRFLFMKNKVRMIVDCHAKHVKVLQDEKLPFDLTLCGSTLRAPHSCHSQYMANMDSIASLVMAVIVNDND--EDGD-SSDAVQPQKRKRLWGLVVCHN
Tipuana A       QASRFLFMKNKVRMIVDCHAKHVKILQNEKLPFDLTLCGSTFRAPHSCHLQYMANMDSIASLVMAVVVNDSD--EDGD-SSDAVQPQKRKRLWGLVVCHN
Xeroderris A    QAARFLFMKNKVRIIVDCHAKHVKVLQDEKLPLDLTLCGSTLRAPHSCHLQYMANMDSIASLFMAVVVNDNE--EDGD-SSDAVQPQKRKRLWGLVVCHN
Millettiar A    QAARFLFMKNKVRMIVDCHAKHVKVLQDEKLPFDLTLCGSTLRAPHSCHLQYMANMDSIASLVMAVVVNDNE--EDGD-SADAVQPQKRKRLWGLVVCHN
Millettiad A    QAARFLFMKNKVRMIVDCHAKHVKVLQDEKLPFDLTLCGSTLRAPHSCHLQYMANMDSIASLVMAVVVNDNE--EDGD-SSDAVQPQKRKRLWGLVVCHN
Lonchocarpus A  QAARFLFMKNKVRMIVDCHAKHVKVLQDEKIPFDLTLCGSTLRAPHSCHLQYMANMDSIASLVMAVVVNDNE--EDGD-SSDAVQPQKRKRLWGLVVCHN
Derris A        QAARFLFMKNKVRMIVDCHAKHVKVLQDEKIPFDLTLCGSTLRAPHSCHLQYMANMDSIASLVMAVVVNDNE--EDGD-SSDAVQPQKRKRLWGLVVCHN
Caragana A      QAARFLFMKEQDPYIVDCHAKHVKVLQDEKLPFDLTLCGSTLRAPHSCHLQYMANMNSIASLVMAVVVNDSD--EDED-SSDAVHPQKRKRLWGLVVCHN
Pisum A         QAARFLFMKNKVRMIVDCNAKHVKVLQDEKLPFDLTLCGSTLRAPHSCHLQYMANMDSIASLVMAVVVNDSD--EDGD-SADAVLPQKKRLWGLVVCHN
Lathyrus A      QAARFLFMKNKVRMIVDCHAKHVKVLQDEKLPFDLTLCGSTLRAPHSCHLQYMANMDSIASLVMAVVVNDSD--EDGD-SRDAVLPQKKRLWGLVVCHN
Wisteria A      QAARFLFMKEQGAYIVDCHAKHVKVLQDEKLPFDLTLCGSTLRAPHSCHLQYMANMDSIASLVMAVVVNDSD--EDGD-SPDAVHPQKRKRLWGLVVCHN
Clianthus A     TRSRFLFMKNKVRMIVDCHAKHVKVLQDEKLPFDLTLCGSTLRAPHSCHLQYMANMNSIASLVMAVVVNDSD--EDGDNSSDAVQPQKKKRLWGLVVCHN

Sesbanias A'    QATRFLFMKNKVRIIVDCSAKHVKVIQDKNIPFDLTLCGSTLRAPHSCHLQYMENMKASASLVMAVVVNDSN--EDGD-SSDAVQPQKRKRLWGLVVCHN
Sesbaniav A'    QATRFLFMKNKVRMIVECCAKHVKVIQDKKNPFDLTLCGSTLRAPHSCHLQYMENMKASASLVMAVVINDSN--EDGD-SSDAVQPQKRKRLWGLVVCHH
Poitea A'       QATRFLFMKNKVRMIVDCSANHVKVLQDKNIPFDLTFCGSTLRAPHSCHLQYMKNMNASASLVMAVVVNDSD--EDGD-SFDVVQLQKRRRLWGLDVCHH
Hybosema A'     QATRFLFMKNKVRMIVDCSANHVKVPQDKNIPFDLTFCGSTLRAPHSCHLQYMKNMNASASLVMAIVVNDSD--EDGD-SSDAVQLQNRRRLWGLVVCHH
Dalbergia A'    QATRFLFMKNRVRMIVDCRAKHVKVLQDLKIPFDLTLCGSTLRAAHSCHLQYMENMNSIASLVMAVVVNDDD--DDGD-GSDVVQPQKRKRLWGLVVCHN
Tipuana A'      QATRFLLMKNKVRMIVDCRAKHVKVLQDPKIPYDLTLCGSTLRAAHSCHLQYMENMNTIASLVMAVVVNDND--EDGD-GSDVVQPQKRKRLWGLVVCHN
Caragana A'     QATLFLFMKNKVHMMVDRRAKHVKVLQDQKIPFDLTLCGSTLRAPHSCHLQYMENMNTSASLVMAVVVNDSN--EDED-SSDDVQLQKRKRLWGLVVCHH
Wisteria A'     QATRFLFMRNKVRMIVDCRAKHVKVLQDKKIPFDLTFCGSTLRAPHSCHLQYMENMNASASLVLAVVVNDND--EDED-SSNDVQPQKRKRLWGLVVCHH
Piscidia A'     QATRFLFMKNKVRMIVDCRAKHVKVLQDEKVPFDLTLCGSTLRAPHSCHLQYMENMNCSASLVMAVVVNDND--EDGD-S-DAVQPQKRKRLWGLVVCHN
Millettiad A'   QATRFLFMKNKVRMIVDCRAKHVKVLQDKKVPFDLTLCGSTLRAPHSCHLQYMENMNCSASLVMAVVVNDND--EDGD-S-DAVQPQKRKRLWGLVVCHN
Lonchocarpus A' QATRFLFMKNKVRMIVDCRAKHVKVLQDKKVPFDLTLCGSTLRAPHSCHLQYMENMNCSASLVMAVVVNDND--EDGD-S-DAVQPQKRKRLWGLVVCHN

Myrospermum E   QAARFLFKQNRVRMICDCYAKPVKVIQSEELRRPLCLVNSTLRSPHGCHTQYMANMGSIASLVMAVTVNGNDTT----------------RLWGLLVCHH
Hebestigma E    QAARFLFKQNRVRMITDCHAKPVKVIQSEELRQPLCLVNSTLRSPHECHTQYMANMGSIASLVMAVIVNRNDTT----------------RLWGLLVCHH
Lennea E        QAARFLFKQNRVRMITDCNAKPVKVIQSEELKQPLCLVNSTLRSPHECHTQYMANMGSIASLVMAVIVNENDTT----------------RLWGLLVCHH
Dalbergia E     QASRFLFRQNRVRMICDCHAKPVKVFQSEELRQPLCLVNSTLRSPHGCHTQYMANMGSIASLVMAIIVNGKHTT----------------RLWGLLVCHH
Dalbergiella E  QAARFLFKQNRVRMICDCHAKPVKVIQSGELRQPLCLVNSTLRSPHGCHTQYMANMGSISSLVMAVVVNGNDTT----------------RLWGLLVCHH
Kunstleria E    QASRFLFKQNRVRMICDCHAKPVKVIQSEELRQPLCLVNSTLRSPHGCHTQYMANMGSIASLVMAVTVNGNDTR----------------RLWGLLVCHH
Xeroderris E    QACRFLFKQNRVRMICDCHAKPVKVIQSEELRQPLCLVNSTLRSPHGCHTQYMANMGSVASLVMAVIVNGNDTT----------------RLWGLLVCHH
Lonchocarpus E  HASRFLFKQNRVRMICDCDAKPVKVIQSEELRQPLCLVNSTLRSPHVCHTQYMANMGSTASLVLAIIVNGNDTT----------------RLWGLLVCHH
Derris E        QASRFLFKQNRVRMICDCHAKPVKVIQSEELRQPLCLVNSTLRSPHVCHTQYMANMGSIASLVMAIIVNGNDTT----------------RLWGLLVCHH
Millettiar E    QASRFLFKQNRVRMICDCHAKPVKVIQSEELRQPLCLVNSTLRSPHVCHTQYMANMGSIASLVMAIIVNGNDKT----------------RLWGLLVCHH
Millettiad E    QASRFLFKQNRVRMICDCNAKPVKVIQSEELRQPLCLVNSTLRSPHVCHTQYMANMGSLASLVMAIIVNGNDTT----------------RLWGLLVCHH
Piscidia E      QASRFLFRQNRVRMICDCHAKPVKVIQSEELRQPLCLVNSTLRSPHVCHTQYMSNMGSIASLVMAIIVNGNDKR----------------RLWGLLVCHH
Wisteria E      QASRFLFKQNRVRMICDCHAKPVKVIQSEELRQPLCLVNSTLRSPHDCHTHYMANMGSIASLVMAVIVNANDAP----------------RLWGLLVCHH
```

APPENDIX 4. Aligned amino acid sequences inferred from nucleotide sequences amplified from the Fabaceae, with 6 C-terminal residues deleted. Amino acids are numbered relative to those in Appendix 1. Orthologous *Arabidopsis PHY* loci are indicated to the right of the name of the genus.

APPENDIX 5. Relative rate tests (Wu & Li, 1985) to detect rate asymmetry. * $P < 0.05$; ** $P < 0.01$. d_{13} and d_{23} are the number of nonsynonymous (or synonymous in legume comparisons) substitutions per site between species 1 and 3, and species 2 and 3, respectively; under the null hypothesis $d_{13} = d_{23}$. SE is standard error.

Species 1	Species 2	Species 3 (reference)	$d_{13} - d_{23} \pm$ SE
		Chromophore region only (330–594 bp) compared	
ArabidopsisA	CucurbitaA	Selaginella	0.0082 ± 0.0194
ArabidopsisA	SolanumA	Selaginella	0.0024 ± 0.0415
ArabidopsisA	OryzaA	Selaginella	−0.0033 ± 0.0416
ArabidopsisA	ArabidopsisB	Selaginella	0.0451 ± 0.0395
ArabidopsisA	ArabidopsisC	Selaginella	0.0761 ± 0.0379
ArabidopsisA	ArabidopsisD	Selaginella	0.0537 ± 0.0390
ArabidopsisA	ArabidopsisE	Selaginella	0.0709 ± 0.0387
SolanumA	SolanumB	Selginella	0.0847 ± 0.0373*
OryzaA	OryzaB	Selaginella	0.0639 ± 0.0387
AvenaA	ZeaA	Selaginella	−0.0081 ± 0.0161
MyrospermumA	HebestigmaA	PisumA	0.0661 ± 0.1679
MilletiaA	SesbaniaA	MyrospermumA	0.1337 ± 0.1350
PisumA	HebestigmaA	MyrospermumA	0.2012 ± 0.1466
HebestigmaE	MilletiaE	MyrospermumE	−0.0536 ± 0.1002
		N-Terminal encoding sequence (2400 bp) compared	
ArabidopsisA	CucurbitaA	Selaginella	0.0064 ± 0.0216
ArabidopsisA	SolanumA	Selaginella	−0.0107 ± 0.0220
ArabidopsisA	OryzaA	Selaginella	0.0053 ± 0.0216
ArabidopsisA	ArabidopsisB	Selaginella	0.0272 ± 0.0253
ArabidopsisA	ArabidopsisC	Selaginella	−0.0115 ± 0.0214
ArabidopsisA	ArabidopsisD	Selaginella	0.0182 ± 0.0213
ArabidopsisA	ArabidopsisE	Selaginella	−0.0228 ± 0.0225
ArabidopsisB	ArabidopsisC	Selaginella	−0.0387 ± 0.0221
ArabidopsisB	ArabidopsisE	Selaginella	−0.0500 ± 0.0218*
ArabidopsisB	SolanumB	Selaginella	0.0346 ± 0.0195
ArabidopsisB	OryzaB	Selaginella	−0.0008 ± 0.0204
SolanumA	SolanumB	Selaginella	0.0725 ± 0.0205**
OryzaA	OryzaB	Selaginella	0.0211 ± 0.0210
		Full-length coding sequence (3384 bp) compared	
ArabidopsisA	CucurbitaA	PisumA	−0.0112 ± 0.0111
ArabidopsisA	SolanumA	Selaginella	−0.0145 ± 0.0199
ArabidopsisA	OryzaA	Selaginella	−0.0226 ± 0.0202
ArabidopsisA	ArabidopsisB	Selaginella	0.0476 ± 0.0187*
ArabidopsisA	ArabidopsisC	Selaginella	−0.0346 ± 0.0206
ArabidopsisA	ArabidopsisD	Selaginella	0.0339 ± 0.0190
ArabidopsisA	ArabidopsisE	Selaginella	−0.0273 ± 0.0204
ArabidopsisE	ArabidopsisB	Selaginella	0.0749 ± 0.0194**
ArabidopsisE	ArabidopsisD	Selaginella	0.0612 ± 0.0197**
ArabidopsisC	ArabidopsisB	Selaginella	0.0822 ± 0.0194**
ArabidopsisC	ArabidopsisD	Selaginella	0.0685 ± 0.0197**
ArabidopsisB	SolanumB	Selaginella	0.0326 ± 0.0234
ArabidopsisB	OryzaB	Selaginella	−0.0085 ± 0.0179
SolanumA	SolanumB	Selaginella	0.0947 ± 0.0183**
OryzaA	OryzaB	Selaginella	0.0617 ± 0.0194**